

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig–Maximilians–Universität München

**Structural and functional
characterization of
Rubisco assembly chaperones**

Thomas Hauser

aus

Tübingen, Deutschland

2016

Erklärung

Diese Dissertation wurde im Sinne von §7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. F. Ulrich Hartl betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 03.02.2016

Thomas Hauser

Dissertation eingereicht am: 25.02.2016

1. Gutachter: Prof. Dr. F. Ulrich Hartl

2. Gutachter: Prof. Dr. Jörg Nickelsen

Mündliche Prüfung am 28.04.2016

Acknowledgements

First of all, I am very thankful to Prof. Dr. F. Ulrich Hartl and Dr. Manajit Hayer-Hartl for giving me the opportunity to conduct my PhD in their department at the Max Planck Institute of Biochemistry. This work has benefited greatly from their scientific expertise and experience together with their intellectual ability to tackle fundamental scientific questions comprehensively. Their way of approaching complex projects has shaped my idea on how to perform science.

I am very grateful to Dr. Andreas Bracher for giving crucial input and collaborating on many aspects on my work conducted in the department. His extensive crystallographic expertise was of great importance during my time as a PhD student.

Furthermore, I want to thank Oliver Müller-Cajar for introducing me into the field of Rubisco and supporting me with help and suggestions in the beginning of my PhD. His enthusiasm about conducting science was of great importance to me and influenced my motivation to work and live science on a day-by-day lab basis enormously.

Thanks to all the people in the Hartl lab, especially the “Rubisco group”, represented by Oliver Müller-Cajar, Mathias Stotz, Candace Tsai, Cuimin Liu and Amanda Windhof for the friendly, sophisticated and open-minded atmosphere. Their comments on my work have been inspiring for my study. It was always a pleasure being with them, in work as well as in private.

I am thankful for the discussions with my thesis advisory committee members Prof. Dr. Jörg Nickelsen and Prof. Dr. Ute Vothknecht and for the support of the IMPRS organisation team Dr. Hans Joerg Schaeffer, Dr. Ingrid Wolf and Maximiliane Reif.

I especially want to thank all collaborators within and outside of the Department. Dr. Petra Wendler and Goran Milicic for performing negative stain electron microscopy, Dr. Javaid Bhat for mass spectrometry, Karina Valer and the MPIB Crystallization facility. Thank you to Evelyn Frey-Royston, Silke Leuze-Bütün and Andrea Obermayr-Rauter for their administrative support as well as Emmanuel Burghardt, Romy Lange, Nadine Wischnewski, Elisabeth Schreil, Albert Ries and especially Anastacia Jungclaus for their technical support.

I want to thank the members of my PhD committee Prof. Dr. Jörg Nickelsen, Prof. Dr. Ute Vothknecht, Dr. Petra Wendler, Dr. Dietmar Martin, and Prof. Dr. Elena Conti for critical evaluation of this thesis.

Of great importance during my stay in Munich were Amit Gupta, Timm Hassemer, Leonhardt Popilka, Tobias Neudegger, David Balchin, Christian Löw, and Mark Hipp. I am deeply obliged for countless memorable moments; they were colleagues and became friends.

A special thank goes to my family. I am deeply grateful for their unlimited support since the beginning of my studies until the end of my PhD.

Finally, I want to express my strongest gratitude to my darling Tanja for her invaluable support and love. You are most important to me and make my life meaningful.

Table of content

| | |
|---|----|
| Summary | 1 |
| 1. Introduction..... | 3 |
| 1.1 Protein Folding..... | 3 |
| 1.2 Molecular Chaperones..... | 4 |
| 1.2.1 The chaperone concept | 4 |
| 1.2.2 The Hsp70 system | 7 |
| 1.2.3 Chaperonins..... | 10 |
| 1.2.4 Assembly chaperones..... | 13 |
| 1.2.4.1 Nucleosome assembly | 13 |
| 1.2.4.2 Proteasome assembly | 14 |
| 1.2.4.3 Ribosome assembly | 14 |
| 1.2.4.4 Pilus assembly..... | 15 |
| 1.3 Photosynthesis | 16 |
| 1.3.1 Light-dependent reactions..... | 17 |
| 1.3.2 Light-independent reactions / Calvin Benson Bassham cycle | 19 |
| 1.3.3 Photorespiration | 21 |
| 1.3.4 Carbon concentrating mechanisms..... | 22 |
| 1.3.4.1 The Carboxysome | 22 |
| 1.3.4.2 The pyrenoid..... | 24 |
| 1.3.4.3 C4 and crassulacean acid metabolism (CAM) photosynthesis..... | 24 |
| 1.4 Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) | 25 |
| 1.4.1 Structure of Rubisco..... | 25 |
| 1.4.2 Catalytic mechanism and regulation of Rubisco activity | 29 |
| 1.4.3 Rubisco: synthesis, folding and assembly | 33 |
| 1.4.4 Role of auxiliary proteins in Rubisco biogenesis and function | 38 |
| 1.5 Aim of the study | 50 |
| 2. Publications | 51 |
| 2.1 Structure and mechanism of the Rubisco-assembly chaperone Raf1 | 51 |

Table of content

| | |
|---|-----|
| 2.2 Structural Analysis of the Rubisco-Assembly Chaperone RbcX-II from <i>Chlamydomonas reinhardtii</i> | 81 |
| 3. Additional unpublished results..... | 99 |
| 3.1 Analysis of Raf1-RbcL interaction by RbcL peptide membrane | 99 |
| 3.2 Raf1-RbcL binding is not affected by Mg ²⁺ , ATP and RuBP | 103 |
| 3.3 RbcS displaces Raf1 from RbcL ₈ at equimolar concentrations | 105 |
| 3.4 Interaction of Raf1 with RbcS | 106 |
| 3.5 Interaction of Raf1, Raf2 and RbcS | 108 |
| 3.6 <i>In vitro</i> reconstitution of plant Rubisco | 111 |
| 3.7 Analysis of eukaryotic RbcX-RbcL interaction | 113 |
| 3.8 Functional analysis of eukaryotic RbcX | 116 |
| 3.9 A potential role for RbcX in Rubisco assembly in <i>Chlamydomonas reinhardtii</i> | 117 |
| 3.10 Interplay of Raf1 and RbcX in Rubisco assembly..... | 119 |
| 4. Discussion | 122 |
| 4.1 A new model for Raf1-mediated Rubisco assembly | 123 |
| 4.2 The role of eukaryotic RbcX in Rubisco assembly | 123 |
| 4.3 Functional interplay between Raf1 and RbcX in Rubisco assembly..... | 125 |
| 4.4 Plant Rubisco <i>in vitro</i> reconstitution..... | 128 |
| 5. Outlook | 131 |
| 6. Material and Methods..... | 132 |
| 6.1 Materials | 132 |
| 6.1.1 Chemicals | 132 |
| 6.1.2 Reagent and purification kits..... | 132 |
| 6.1.3 Enzymes, proteins and antibodies | 133 |
| 6.1.4 Instruments | 133 |
| 6.1.5 Strains | 134 |
| 6.1.6 Media | 135 |
| 6.1.7 Plasmids and oligonucleotides..... | 135 |
| 6.2 Molecular biological methods | 136 |
| 6.2.1 DNA analytical methods..... | 136 |

| | |
|---|-----|
| 6.2.2 PreparaAion and transformation of <i>E.coli</i> cells | 136 |
| 6.2.3 Purification of plasmid DNA and DNA-fragments..... | 136 |
| 6.2.4 PCR (polymerase chain reaction) | 137 |
| 6.2.5 Restriction digest and ligation | 138 |
| 6.2.6 Cloning strategies | 138 |
| 6.3 Protein biochemical and biophysical methods | 139 |
| 6.3.1 Protein quantification..... | 139 |
| 6.3.2 SDS-PAGE..... | 139 |
| 6.3.3 Gradient Native PAGE | 140 |
| 6.3.4 Coomassie blue staining of polyacrylamide gels | 141 |
| 6.3.5 Western blotting and immunodetection | 141 |
| 6.3.6 Protein expression and purification | 142 |
| 6.3.6.1 AtCpn10..... | 142 |
| 6.3.6.2 AtCpn20..... | 142 |
| 6.3.6.3 AtCpn60 $\alpha_7\beta_7$ / AtCpn60 β_7 | 143 |
| 6.3.6.4 AtRaf2 / Syn7002Raf2 / FLAG-AtRbcXI / FLAG-AtRbcXII / CrRbcX-IIb..... | 143 |
| 6.3.6.5 AnaCA-RbcX | 143 |
| 6.3.6.6 Prokaryotic and Eukaryotic RbcL | 144 |
| 6.3.6.7 Prokaryotic and Eukaryotic RbcS | 144 |
| 6.3.7 Co-immunoprecipitation | 145 |
| 6.3.8 Peptide binding screen..... | 145 |
| 6.3.9 Transformation and selection of <i>Chlamydomonas reinhardtii</i> | 146 |
| 6.3.10 <i>Chlamydomonas</i> total RNA isolation..... | 146 |
| 6.3.11 <i>Chlamydomonas</i> total protein analysis | 146 |
| 7. References | 147 |
| 8. Appendices..... | 161 |
| 8.1 Oligonucleotides | 161 |
| 8.2 Plasmids | 162 |
| 8.3 Abbreviations | 163 |
| 8.4 Curriculum Vitae | 166 |

Summary

Life on earth is almost exclusively dependent on the ability of photosynthetic organisms to sequester inorganic carbon dioxide (CO₂) of the atmosphere into organic carbon of the biosphere via the Calvin-Benson-Bassham (CBB) pathway. The primary photosynthetic CO₂ fixation reaction, the binding of CO₂ to the acceptor-molecule ribulose-1,5-bisphosphate (RuBP) to form two molecules of 3-phosphoglycerate, is catalyzed by the enzyme RuBP carboxylase/oxygenase (Rubisco). It is the major enzyme assimilating CO₂ into the biosphere and thought to be the most abundant protein in nature. At the same time Rubisco is an inefficient catalyst, and it also catalyzes a wasteful side-reaction with atmospheric oxygen. Downstream processing of the oxygenation product necessitates an energetically costly pathway, called photorespiration, during which CO₂ is released. As a consequence of its shortcomings the enzyme is often the rate-limiting step of photosynthesis. These inadequacies make Rubisco a prime target for increasing agricultural productivity.

An enzyme becomes only active when it reaches its final tertiary or quaternary structure. Cells have evolved a complex machinery called molecular chaperones that prevent protein aggregation and promote the proper folding/assembly of a polypeptide chain. The correct folding of Rubisco is dependent on molecular chaperones called chaperonins. Form I Rubisco is a hexadecameric protein complex consisting of an octameric core of large subunits that is capped on both sides by four small subunits. For some cyanobacterial Rubiscos this folding is sufficient to allow productive assembly of the holoenzyme. Nevertheless, recent findings show that additional factors are involved in Rubisco assembly. One of these factors is called RbcX and was found to be an assembly chaperone for cyanobacterial Rubisco. Furthermore, a potential new Rubisco assembly factor was found in maize, called Rubisco accumulation factor 1 (Raf1). Raf1 knock-out plants were seedling lethal and exhibited greatly reduced Rubisco levels. However, the potential role of Raf1 in Rubisco assembly and the pathway of folding and assembly of Rubisco in general is still not fully understood, in particular for higher plant Rubisco.

In the present study, the structure and mechanism of two assembly chaperones of Rubisco, Raf1 and RbcX, were investigated. The role of Raf1 in Rubisco assembly was elucidated by analyzing cyanobacterial and plant Raf1 with a vast array of biochemical and biophysical techniques. Raf1 is a dimeric protein. The subunits have a two-domain structure. The crystal structures of two separate domains of *Arabidopsis thaliana* (At) Raf1 were solved at resolutions of 1.95 Å and 2.6–2.8 Å, respectively. The oligomeric state of Raf1 proteins was investigated by size exclusion chromatography connected to multi angle light scattering (SEC-MALS) and native mass spectrometry (MS). Both cyanobacterial and plant Raf1 are dimeric with an N-

terminal domain that is connected via a flexible linker to the C-terminal dimerization domain. Both Raf1 proteins were able to promote assembly of cyanobacterial Rubisco in an *in vitro* reconstitution system. The homologous cyanobacterial system resulted in very high yields of active Rubisco (>90%), showing the great efficiency of Raf1 mediated Rubisco assembly. Two distinct oligomeric complex assemblies in the assembly reaction could be identified via native PAGE immunoblot analyses as well as SEC-MALS and native MS. Furthermore, a structure-guided mutational analysis of Raf1 conserved residues in both domains was performed and residues crucial for Raf1 function were identified. A new model of Raf1 mediated Rubisco assembly could be proposed by analyzing the Raf1-Rubisco oligomeric complex with negative stain electron microscopy. The final model was validated by determining Raf1-Rubisco interaction sites using chemical crosslinking in combination with mass spectrometry. Taken together, Raf1 acts downstream of chaperonin-assisted Rubisco large subunit (RbcL) folding by stabilizing RbcL antiparallel dimers for assembly into RbcL₈ complexes with four Raf1 dimers bound. Raf1 displacement by Rubisco small subunit (RbcS) results in holoenzyme formation.

In the second part of this thesis, the role of eukaryotic RbcX proteins in Rubisco assembly was investigated. Eukaryotes have two distinct homologs of RbcX, RbcX-I and RbcX-II. Both, plant and algal RbcX proteins were found to promote cyanobacterial Rubisco assembly in an *in vitro* reconstitution system. Mutation of a conserved residue important for Rubisco assembly in cyanobacterial RbcX also abolished assembly by eukaryotic RbcX, underlining functional similarities among RbcX proteins from different species. The crystal structure of *Chlamydomonas reinhardtii* (Cr) RbcX was solved at a resolution of 2.0 Å. RbcX forms an arc-shaped dimer with a central hydrophobic cleft for binding the C-terminal sequence of RbcL. Structural analysis of a fusion protein of CrRbcX and the C-terminal peptide of RbcL suggests that the peptide binding mode of CrRbcX may differ from that of cyanobacterial RbcX. RbcX homologs appear to have adapted to their cognate Rubisco clients as a result of co-evolution. Preliminary analysis of RbcX in *Chlamydomonas* indicated that the protein functions as a Rubisco assembly chaperone *in vivo*. Therefore, RbcX was silenced using RNAi in *Chlamydomonas* which resulted in a photosynthetic growth defect in several transformants when grown under light. RbcX mRNA levels were highly decreased in these transformants which resulted in a concomitant decrease of Rubisco large subunit levels. Biochemical and structural analysis from both independent studies in this thesis show that Raf1 and RbcX fulfill similar roles in Rubisco assembly, thus suggesting that functionally redundant factors ensure efficient Rubisco biogenesis.

1. Introduction

Proteins are abundant biological macromolecules consisting of amino acids linked together by peptide bonds. They are found ubiquitously in all domains of life and are essential for almost all biological functions including metabolism, signaling and transport. The function of a protein is determined by its three-dimensional structure, which ultimately is encoded in the primary sequence of the amino acids.

1.1 Protein Folding

Newly synthesized proteins must fold into a unique three-dimensional structure to perform their specific cellular functions. Based on refolding experiments with Ribonuclease A, Christian Anfinsen was the first to postulate that the information of the final folded structure of a protein is determined by the primary sequence of amino acids¹⁻⁴. The main driving force of protein folding in solution is the burial of hydrophobic side chains in the interior of the protein and the entropy gain of solvent molecules (water) released during this process. Folding reactions are highly complex and heterogenous, relying on the cooperation of further multiple weak, noncovalent interactions. These include salt bridges between opposite charges, hydrogen bonds and van der Waals contacts between atoms. Thermodynamically, the native state of a protein represents the most stable state of the system being energetically lower than the unfolded state⁴. However, since proteins usually require some degree of flexibility for their biological function, most protein structures represent a compromise between thermodynamic stability and the conformational flexibility⁵. Consequently, proteins are often marginally stable in their physiological environment and thus susceptible to misfolding and aggregation. Protein folding occurs rapidly (microseconds to seconds) demonstrating the existence of directed and kinetically driven folding pathways, rather than sampling all possible conformations before reaching their native state in a biologically irrelevant timescale⁶. Polypeptide chains are thought to explore funnel-shaped potential energy surfaces as they progress toward the native structure along several downhill paths rather than a single defined pathway (Fig.1). At the beginning of folding events, local amino acid sequences initiate folding, followed by the rapid formation of transient, partially folded and more compact intermediates. Subsequently, higher structure elements are formed until the native state is accomplished. However, energy landscapes can be "rugged" and therefore have local energy minima that are associated with kinetically trapped intermediates. The "ruggedness" of the energy landscape determines the speed of folding, since more local minima lead to more kinetically trapped intermediates and therefore slower folding. Especially complex architectures often fold very slow (timescale of min to h) or even fail to reach their native state in vitro, resulting in partially folded intermediates. Furthermore, misfolded states can arise along the folding pathway. These are

characterized by the presence of nonnative interactions that must be resolved prior to correct folding. Partially folded or misfolded states typically expose hydrophobic amino acid residues and regions of unstructured polypeptide backbone to the solvent, the features that mediate aggregation in a concentration-dependent manner⁷. Although aggregation primarily leads to amorphous structures largely driven by hydrophobic forces, it may also lead to the formation of amyloid-like fibrils which are associated with protein folding diseases.

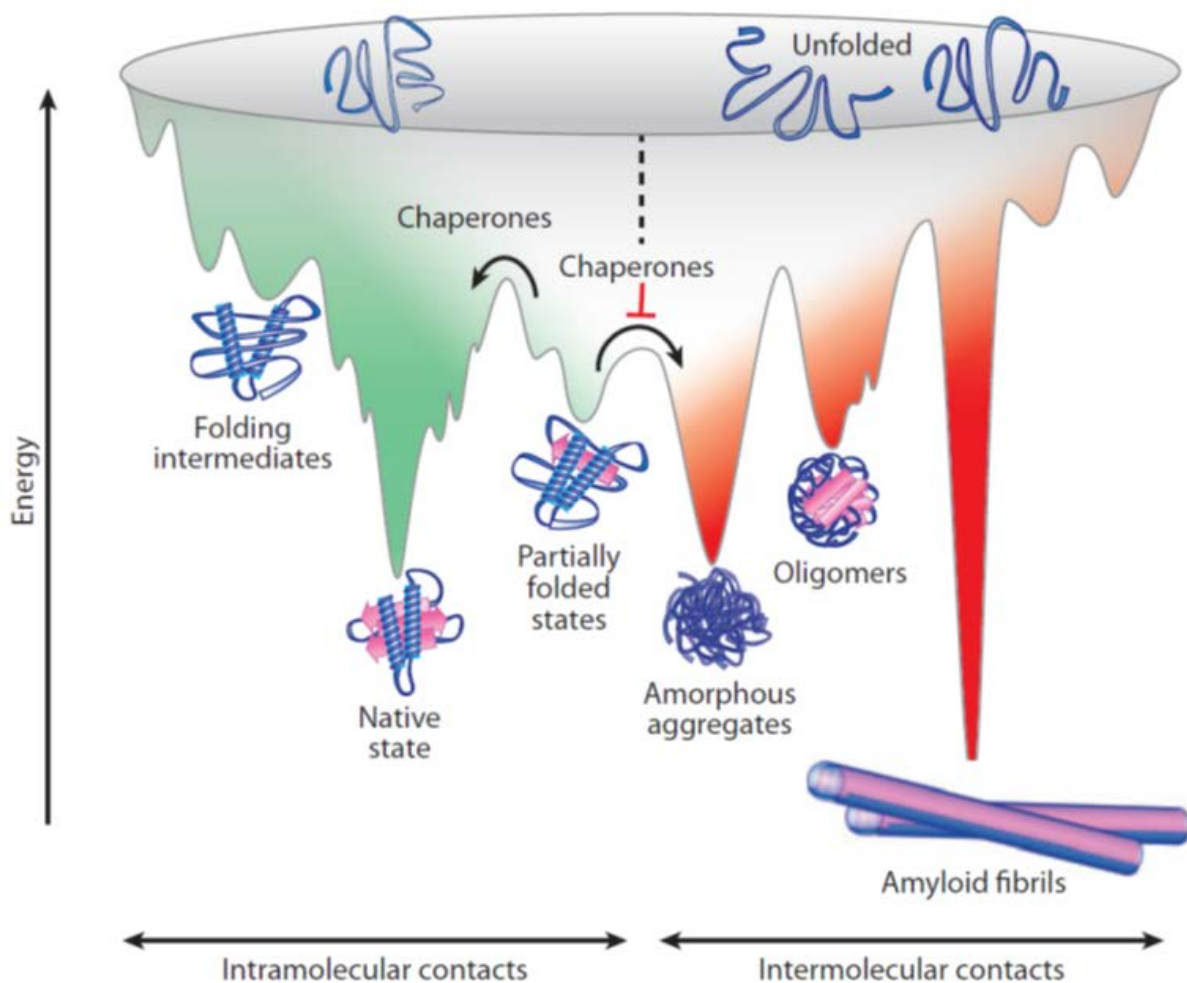


Fig.1: Energy landscape in protein folding and aggregation. Left: energetically favorable intramolecular interactions (green) result in folding to the native state. Energetically favorable but nonnative conformations result in kinetically trapped states. Right: intermolecular interactions (red) leading to aggregation (amorphous aggregates, β -sheet-rich oligomers, and amyloid fibrils; reproduced from Kim *et al.*⁷).

1.2 Molecular Chaperones

1.2.1 The chaperone concept

The native fold of a protein is encoded in its amino acid sequence⁸. To become fully active, newly synthesized proteins must fold to unique three-dimensional structures. Spontaneous refolding *in vitro* is generally efficient for small, single domain proteins that bury exposed

Introduction

hydrophobic amino acid residues rapidly upon initiation of folding. In contrast, larger proteins with more complex domains or composed of multiple domains often refold inefficiently, owing to the formation of partially folded intermediates, including misfolded states, that tend to aggregate. Furthermore, compared to the *in vitro* refolding experiments, the situation in cells is considerably more complex, as they are confronted with the task of folding thousands of different polypeptides into a wide range of conformations. Moreover, the cellular environment is extremely crowded with high concentrations of proteins, nucleic acids and other macromolecules up to 300-400 mg/ml⁹. The resulting excluded volume effect, or macromolecular crowding, has several consequences for the physical properties of the intracellular environment, and one of the most important is that intermolecular interactions become strongly favored. For an unfolded polypeptide, this means that non-productive aggregation with other unfolded species, which is mediated by exposed hydrophobic residues and unstructured chain segments, can compete effectively with its folding to the native state¹⁰. These exposed hydrophobic surfaces occur on nascent chains, on mature proteins unfolded by environmental stresses, and on folded proteins in near-native and native conformations¹¹. The aggregation process irreversibly removes proteins from their productive folding pathways, and must be prevented *in vivo* by molecular chaperones, which assist proteins to fold on many stages along their pathway to their final active three-dimensional structure. In the cytosol of prokaryotic and eukaryotic cells, molecular chaperones of different structural classes form a network of pathways that can handle substrate polypeptides from the point of initial synthesis on ribosomes to the final stages of folding^{12,13}. Molecular chaperones are defined as a large and diverse group of proteins that share the property of assisting the non-covalent folding and unfolding and the assembly and disassembly of other macromolecular structures, but are not permanent components of these structures when they are performing their normal biological functions¹⁴. Some, but not all, chaperones are also stress or heat shock proteins (Hsps), because the requirement for chaperone function increases under conditions of stress that cause proteins to unfold and to misassemble. Members of the various groups of chaperones were initially named according to their molecular weight in kilo Dalton (kDa): Hsp40s, Hsp60s, Hsp70s, Hsp90s, Hsp100s, and the small Hsps. The assistance of protein folding by molecular chaperones begins with the appearance of the newly-synthesized polypeptide at the ribosome exit tunnel. Because the formation of a stable tertiary structure is a cooperative process at the level of protein domains (50 to 300 amino acid residues), an average domain can completely fold only when its entire sequence has emerged from the ribosome. As a consequence, many nascent chains expose non-native features for a considerable length of time and are prone to aggregation. This tendency to aggregate is thought to be greatly increased by the close proximity of nascent chains of the same type in polyribosome complexes, thus leading to the requirement for chaperones to maintain nascent chains in a non-aggregated, folding-

competent conformation¹⁵. Contrary to the previous belief that polysomes enhance aggregation by increasing the local concentration of nascent chains, recent studies suggest that the three-dimensional organization of individual ribosomes in polysomes maximizes the distance between nascent chains, thus reducing the probability of unproductive interactions^{16,17}. In the cytosol, chaperones generally recognize hydrophobic residues and/or unstructured backbone regions in their substrates, i.e., structural features typically exposed by non-native proteins but normally buried upon completion of folding. Interactions with chaperones may not only block intermolecular aggregation directly by shielding the sticky surfaces of non-native polypeptides, including unassembled protein subunits, but may also prevent or reverse intramolecular misfolding. Cytosolic chaperones participate in *de novo* folding mainly through two distinct mechanisms. Chaperones, such as trigger factor and the Hsp70s, act by holding nascent and newly synthesized chains in a state competent for folding upon release into the medium. In contrast, the large, cylindrical chaperonin complexes provide physically defined compartments inside which a complete protein or a protein domain can fold while being sequestered from the cytosol⁷. These two classes of chaperones are conserved in all three domains of life and can cooperate in a topologically and timely ordered manner^{18,19} (Fig.2).

Besides their fundamental role in *de novo* protein folding, chaperones are involved in various aspects of proteome maintenance, including assistance in macromolecular complex assembly, protein transport and degradation, aggregate dissociation and refolding of stress-denatured proteins. By maintaining nonnative proteins in a soluble, folding-competent state, chaperones are thought to buffer mutations in proteins that would otherwise preclude their folding, thus broadening the range of mutant proteins subject to Darwinian selection^{20,21}. In the following sections the Hsp60/chaperonin system and the Hsp70 system are discussed as examples for the complex mechanisms of molecular chaperones.

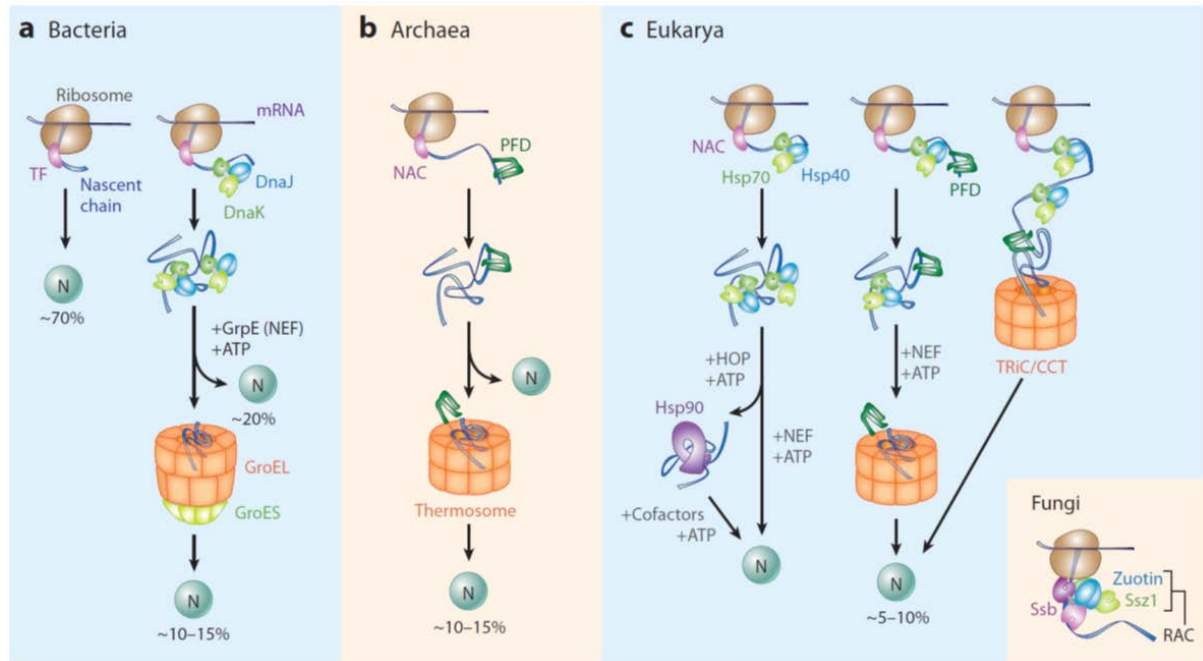


Fig.2: Models for the chaperone-assisted folding of newly synthesized polypeptides in the cytosol. A: Eubacteria. Nascent chains probably interact generally with trigger factor (TF), and most small proteins (70% of total) fold rapidly upon synthesis without further assistance to their native state (N). Longer chains (30% of total) interact subsequently with DnaK and DnaJ and fold upon one or several cycles of ATP-dependent binding and release. About 10 to 15% of chains transit the chaperonin system, GroEL and GroES, for folding. B: Archaea. Only some archaeal species contain DnaK/DnaJ. The existence of a ribosome-bound nascent chain-associated complex (NAC) homolog, as well as the interaction of prefoldin (PFD) with nascent chains, has not yet been confirmed experimentally. C: Eukarya. Like TF, NAC probably interacts generally with nascent chains. The majority of small chains may fold upon ribosome release without further assistance. Other chains reach their native states in a reaction assisted by Hsp70 and Hsp40, and a fraction of these must be transferred to Hsp90 for folding. About 5-10% of chains are co- or posttranslationally passed on to the chaperonin TRiC in a reaction mediated by PFD. Insert: the ribosome-binding chaperone system, the ribosome-associated complex (RAC), in fungi. RAC consists of Ssz1 (a specialized Hsp70) and zuotin (Hsp40) and assists nascent chain folding together with another Hsp70 isoform, Ssb. (reproduced from Kim *et al.*⁷).

1.2.2 The Hsp70 system

The evolutionary conserved Hsp70s have a central role in the cytosolic chaperone network. They are involved in a wide range of protein quality control functions, including de novo protein folding, refolding of stress denatured proteins, protein transport, membrane translocation, and protein degradation. Hsp70 chaperones function with cochaperones of the Hsp40 family (also known as DnaJ proteins or J proteins) and nucleotide exchange factors (NEFs) to assist folding co- or posttranslationally through ATP-regulated cycles of substrate binding and release. Moreover, they mediate polypeptide chain transfer to downstream chaperones. Hsp70 consists of an N-terminal nucleotide-binding domain (NBD) and a C-terminal substrate-binding domain (SBD) connected by a highly conserved hydrophobic linker region (Fig.3a). The N-terminal domain consists of two lobes, each containing two subdomains, with the nucleotide-binding cleft situated in between^{22,23}. The SBD consists of a β -sandwich subdomain and an α -helical lid with the substrate binding site located in the β -sandwich subdomain²² (Fig.3a). The

SBD binds to 5-7-residue peptide segments enriched in hydrophobic amino acids and typically flanked by positively charged residues. Such segments occur on average every 50-100 amino acids in proteins, and the exposure of these fragments correlates with the aggregation propensity of the protein²⁴. The α -helical lid and a conformational change in the β -sandwich domain regulate the affinity state for the peptide in an ATP-dependent manner. In the ATP-bound state, the lid adopts an open conformation, resulting in high on and off rates for the peptide^{25,26}. Hydrolysis of ATP to ADP is strongly accelerated by Hsp40 (>1000-fold), leading to lid closure and stable peptide binding (low on and off rates for the peptide substrate)^{5,27,28} (Fig.3b). Hsp40 (J protein) and NEF cochaperones regulate the Hsp70 reaction cycle^{5,30}. The Hsp40 proteins constitute a large family with more than 40 members in humans²⁹. All of them contain a J domain, which binds to the N-terminal ATPase domain of Hsp70 and the adjacent linker region^{31,32}. Canonical Hsp40s (class I and II) interact directly with unfolded polypeptides and can recruit Hsp70 to protein substrates^{18,30}. Other Hsp40s (class III) are more diverse and combine the J domain with a variety of functional modules^{29,33,34}. After ATP hydrolysis, a nucleotide-exchange factor binds to the Hsp70 ATPase domain, catalyzing ADP-ATP exchange, resulting in lid opening and substrate release. Substrate release allows folding (i.e., the burial of hydrophobic residues) to proceed. Proteins unable to fold rapidly upon dissociation from Hsp70 may rebind, transfer to downstream chaperones (the chaperonin system), or be targeted to the degradation machinery.

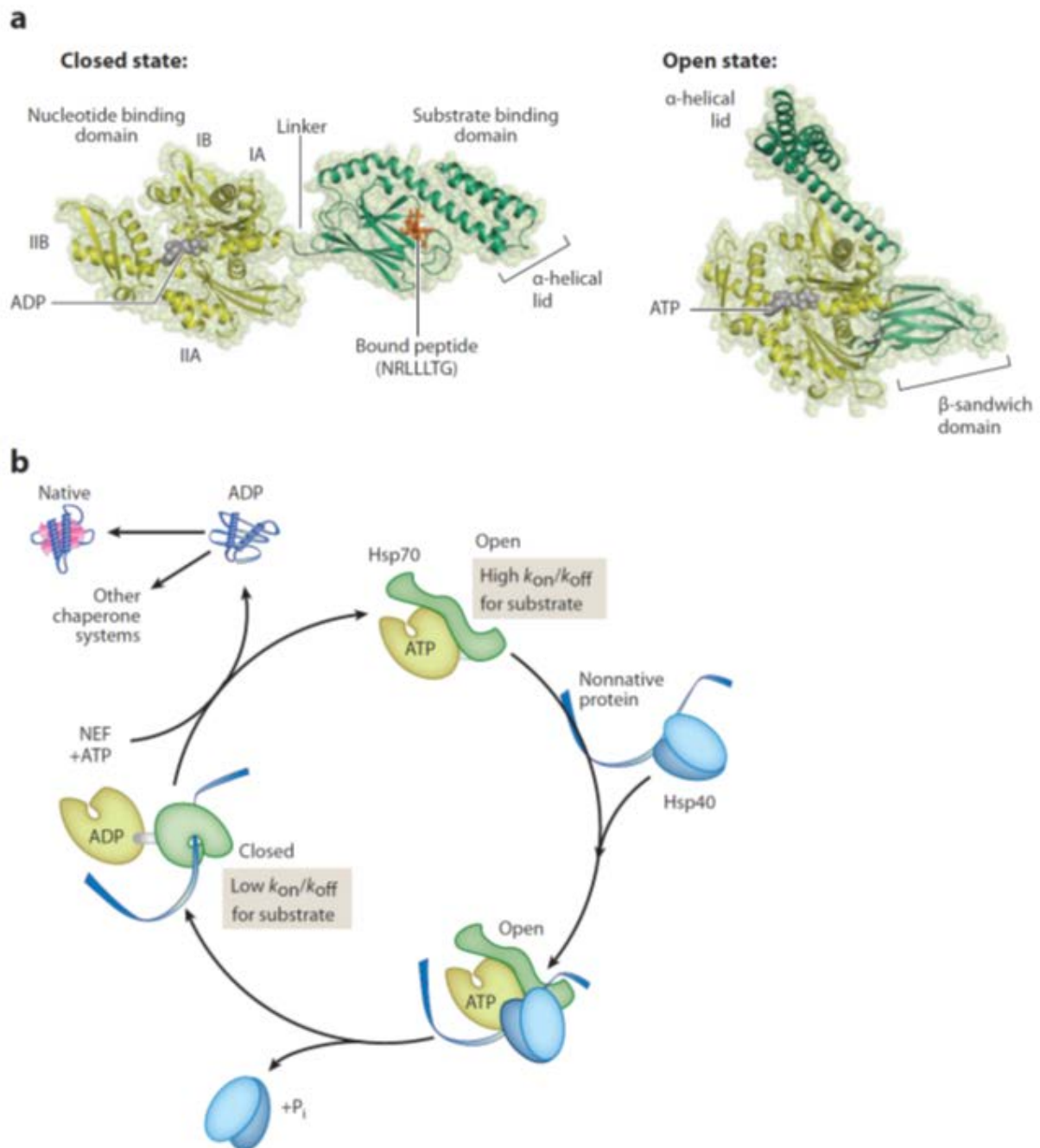


Fig.3: Structure and functional cycle of Hsp70. A, left: Closed state of *E. coli* DnaK (PDB 2KHO), illustrating the ADP-bound NBD separated by a linker from the SBD. The α -helical lid of the SBD is closed over the substrate peptide (NRLLLTG) bound in the pocket of the β -sandwich domain. Right, open state illustrated by the structure of ATP-bound Sse1 (PDB 2QXL). The β -sandwich domain contacts subdomain IA of the NBD, the α -helical lid contacts subdomains IA and IB. B: Hsp70 reaction cycle. ATP binding to the NBD stabilizes the open state of Hsp70, facilitating the binding of substrate protein recruited to Hsp70 by Hsp40 cochaperone. The open state has fast on and off rates for substrate peptide. Hsp40 stimulates ATP hydrolysis on Hsp70, resulting in the closing of the SBD α -helical lid over the bound substrate peptide. The closed state has slow on and off rates for substrate peptide. NEFs stimulate ADP release from the NBD, ATP binding causes substrate release (reproduced from Kim *et al.*⁷).

1.2.3 Chaperonins

Chaperonins are a family of sequence-related proteins of ~60 kDa which form double-ring complexes of ~800-900 kDa that enclose a central cavity in each ring. They occur in two distinct subgroups. The eubacterial GroEL, mitochondrial Hsp60 and plant chaperonin from chloroplasts have seven-membered rings and constitute the group I chaperonins. They cooperate with cofactors of the GroES or Hsp10 (Cpn10) family. Most group II chaperonins have eight-membered rings and are represented by TRiC (TCP-1 containing ring complex; also called chaperonin containing TCP-1 or CCT), the cytosolic chaperonin of eukaryotes, and by the related archaeal chaperonin, known as the thermosome. They are GroES or Hsp10 (Cpn10) independent, possibly because the function of GroES has been incorporated into the chaperonin itself through appropriate structural adaptations³⁵⁻³⁷. In both cases, non-native substrate protein is first captured via hydrophobic contacts with multiple subunits in a ring, and is then displaced into the central cavity where it can fold, while being protected from aggregation with other non-native proteins³⁸. The best studied chaperonin is the group I chaperonin GroEL with its co-chaperonin GroES from *E.coli*. In GroEL, two heptameric rings of 57 kDa subunits are stacked back-to-back and form two cavities of 45 Å width³⁹. Each subunit consists of three domains: the equatorial domain harbors the ATP binding site and forms most of the inter-subunit contacts within and between heptameric rings. It is connected via an intermediate hinge-like domain to the apical domain, which is situated at the rim of the cylinder opening and exposes a number of hydrophobic residues towards the ring cavity for the binding of non-native polypeptide (Fig.4).

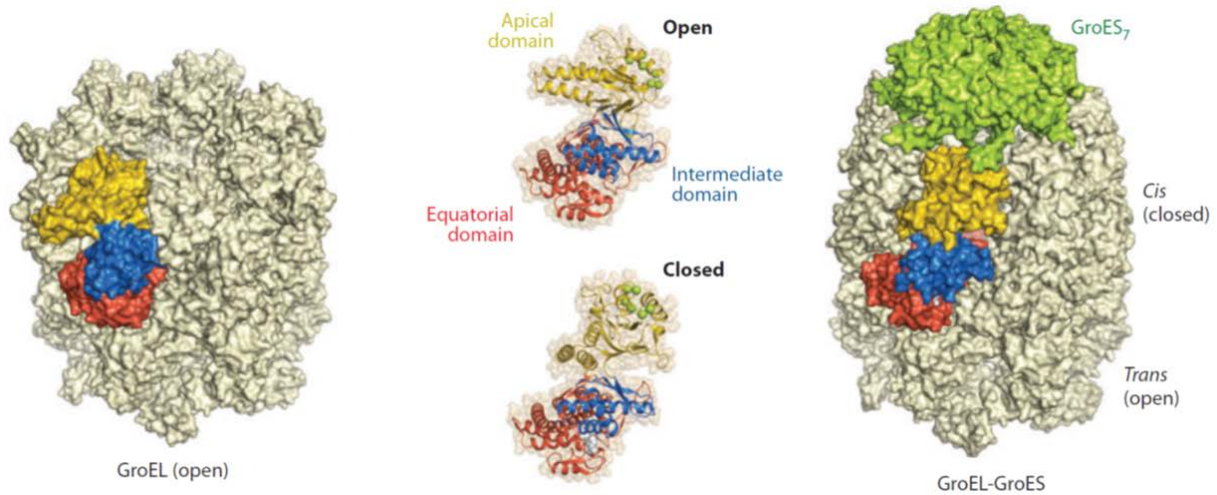


Fig.4: The asymmetric GroEL-GroES-(ADP)₇ complex. The crystal structures of GroEL (PDB 1SS8, left) and the asymmetrical GroEL-(ADP)₇-GroES complex (PDB 1PF9, right) are depicted with GroES shown in green and one subunit of GroEL colored to indicate its domain structure (equatorial nucleotide-binding domain in red; intermediate hinge domain in blue; and the apical substrate and GroES-binding domain in yellow). The conformational differences between the GroEL subunits in the open state (GroEL and *trans* ring of GroEL-GroES complex) and in the closed state (*cis* ring of GroEL-GroES complex) are shown in ribbon representations of single subunits (middle). The green spheres represent hydrophobic residues on helices 8 and 9 of the apical domain that are involved in substrate binding in the open conformation and in GroES binding in the closed state (reproduced from Kim *et al.*⁷).

GroES is a homoheptameric ring of ~10 kDa subunits that cycles on and off the ends of the GroEL cylinder, in a manner regulated by the GroEL ATPase^{40,41}. Hydrophobic sequences of substrate proteins bind to a flexible groove between two amphiphilic helices in the apical domain. The GroES subunits have mobile sequence loops that contact the substrate-binding regions in the apical domains of GroEL and mediate substrate dissociation^{42,43}. The folding cage defined by the GroEL central cavity and the dome-shaped GroES, also termed 'Anfinsen cage'⁴⁴ is characterized by rather hydrophilic walls (in contrast to the cavity of GroEL alone), and is able to enclose proteins up to 50-60 kDa^{44,45}. The binding and release of GroES is regulated by the ATPase activity of GroEL in a 'two-stroke' mechanism. Binding of substrate polypeptide and of 7 ATP to GroEL sequentially (to the free site of GroEL, i.e. the *trans* site) precedes binding of GroES, which closes the folding cage and forms the so-called *cis*-cavity. This results in an enlargement of the cavity from ~85000 Å³ to ~175000 Å³⁴⁶. Folding in the cage can proceed for ~10 s, during which the 7 ATP in the *cis*-ring of GroEL are being hydrolyzed. This hydrolysis is followed by the binding of 7 ATP to the *trans*-ring of GroEL, which sends an allosteric signal to the *cis*-ring that results in the release of GroES. Upon opening of the folding cage, folded protein can exit, whereas folding intermediates that still expose a sufficient amount of hydrophobic surface may be recaptured in preparation for a subsequent round of folding in the GroEL-GroES cavity (Fig.5).

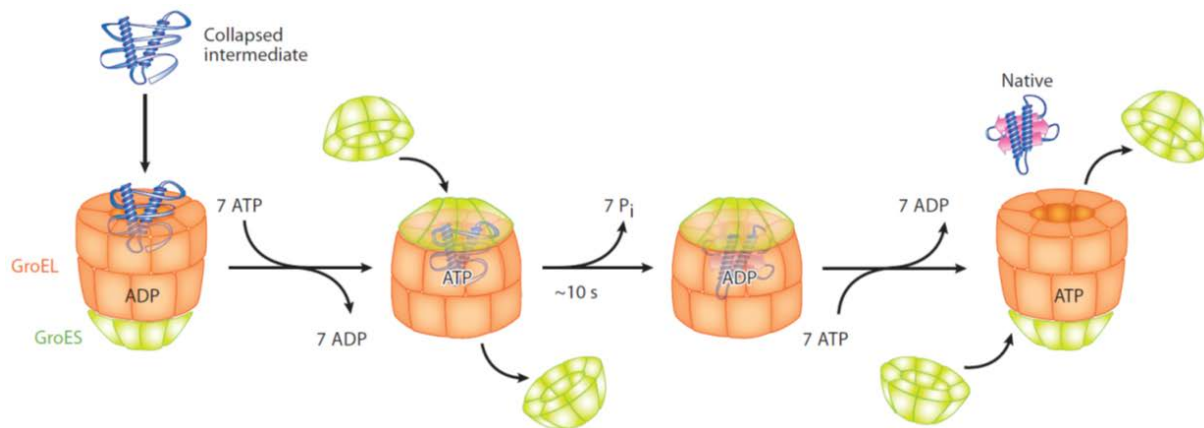


Fig.5: Protein folding in the GroEL-GroES cage. Substrate protein as a collapsed folding intermediate is bound by the open GroEL ring of the asymmetrical GroEL-GroES complex. Binding of ATP to each of the seven GroEL subunits causes a conformational change in the apical domains, which results in the exposure of the GroES binding residues, allowing substrate encapsulation in the *cis* complex. ADP and GroES dissociate from the opposite ring (*trans* ring) together with the previously bound substrate. The newly encapsulated substrate is free to fold in the GroEL cavity during the time needed to hydrolyze the seven ATP molecules bound to the *cis* ring (~10 s). ATP binding followed by GroES binding to the *trans* ring triggers GroES dissociation from the *cis* ring, releasing the substrate protein (reproduced from Kim *et al.*⁷).

About 250 GroEL interacting proteins have been identified in the *E. coli* cytosol, and many of them contain two or more domains with α/β (TIM-barrel) folds^{47,48}. Proteins with such complex topologies typically fold slowly and are aggregation prone, owing to the prolonged exposure of extensive hydrophobic surfaces in their non-native states. Stringent model substrates of GroEL, such as bacterial Rubisco, share this domain topology and fold efficiently only when in the GroEL-GroES cage⁴⁹. There are currently three models for acceleration of folding by the GroEL/GroES system, which differ in whether GroEL/ES solely acts passively by preventing aggregation (passive cage) or additionally promotes the folding process by an active mechanism (active cage and iterative annealing). Although mechanistically distinct, these three models are not mutually exclusive. The central element of the ‘iterative annealing’ hypothesis suggests that the GroES-mediated movement of the apical GroEL domains exerts a stretching force on bound substrate protein, thereby actively unfolding kinetically trapped, misfolded intermediates⁵⁰. This effect, occurring in every chaperonin ATPase cycle, would speed up folding by reducing the half-life of slow folding species and allowing their repartitioning with kinetically more effective folding routes. However, the significance of iterative annealing is unclear considering that a single round of substrate binding and encapsulation (using a single-ring mutant of GroEL that binds GroES stably) results in substrate protein folding with equal efficiency and kinetics as achieved through multiple cycles of binding and release^{9,51,52}. In contrast, ‘cage-mediated annealing’⁵¹ posits that the physical environment of the chaperonin cavity is critical in enhancing folding speed. Growing experimental and theoretical evidence suggests that the GroEL-GroES cage promotes folding by sterically confining folding intermediates^{9,51,53-58}. This model assigns an active role to the

chaperonin cage, as opposed to the view that it functions solely as a passive aggregation prevention device⁵⁹. In addition to steric confinement, mutational analysis demonstrated that the charged residues of the GroEL cavity wall are critical for the observed acceleration^{51,60,61}. According to molecular dynamics simulations, these polar residues accumulate ordered water molecules in their vicinity, thereby generating a local environment in which a substrate protein may bury exposed hydrophobic residues more effectively⁵⁶.

1.2.4 Assembly chaperones

The common perception that molecular chaperones are involved primarily with assisting the folding of newly synthesized and stress-denatured polypeptide chains ignores the fact that this term was coined to describe the function of a protein that was thought to assist the assembly of folded subunits into oligomeric structures and only later was extended to embrace protein folding⁶². There has been a continuing emphasis of chaperone research on the roles of these proteins in protein folding. There is a wealth of experimental data supporting their importance in preventing and reversing the misassembly of newly synthesized polypeptide chains that competes with correct folding (see 1.2.3). The folding of a given polypeptide chain is characterized by the formation of a stable fold specific to the sequence of that chain, whereas assembly is characterized by the association of two or more folded subunits into a biologically functional oligomer. Only recently, the possible role of molecular chaperones in the assembly of oligomeric protein complexes has been increasingly investigated.

1.2.4.1 Nucleosome assembly

The nucleosome is the elemental repeating unit in all eukaryotic chromatin, and consists of 2 copies each of the 4 core histones H2A, H2B, H3, and H4. One tetramer of (H3/H4)₂ and 2 dimers of (H2A/H2B) form the histone octamer, around which 147 bp of DNA are wrapped in 1.7 turns of a tight superhelix⁶³. Thousands of nucleosomes are further compacted into multiple hierarchical levels. Despite the high degree of compaction, chromatin is dynamic and fluidic, and its histone components are exchanged at a relatively high rate⁶⁴. By regulating access to the DNA substrate, the nucleosome is not only a structural unit of the chromosome, but is perhaps the most important regulator of gene expression. Histones are highly basic proteins and require molecular assembly chaperones that prevent them from making improper interactions (either with other proteins, or with DNA), and promote their proper interaction to form chromatin. A number of histone chaperones (e.g., nucleosome assembly protein 1 (NAP1), nucleoplasmin, Asf1, HIR, Spt6, DF31, ACF, REF, Nucleoplasmin/B23, CAF-1, and N1/N2) are involved in chromatin assembly, in conjunction with ATP-dependent chromatin remodeling factors⁶⁵. Many of these activities may also be implicated in modulating chromatin structure by promoting the dynamic exchange of histones in assembled chromatin.

1.2.4.2 Proteasome assembly

In eukaryotes, short-lived proteins are degraded primarily by the ubiquitin-proteasome system⁶⁶. Defects in the system are linked to a variety of human diseases, and proteasomal inhibitors are used to treat several cancers⁶⁷. Most proteasome substrates are first modified by polyubiquitin chains, allowing recognition by the proteasome and degradation of the substrate. The 26S proteasome consists of a proteolytically active 20S proteasome core particle bound at one or both ends by a 19S regulatory particle⁶⁸. Core particles are complexes of 14 different subunits arranged in four heptameric rings. Inside the central chamber, unfolded polypeptide chains are threaded for cleavage into small peptides. In mammals, the two outer rings consist of seven different but related α subunits, and the two inner rings comprise seven different but related β subunits. Eukaryotic core particle assembly initiates with formation of an α ring followed by ordered addition of β subunits to the α ring heteroheptamer⁶⁹. Joining of two half-proteasomes triggers autocatalytic processing of active-subunit propeptides and core particle maturation⁷⁰. Moreover, at least three phylogenetically conserved core particle-specific assembly chaperones (Ump1, PAC proteins) facilitate proteasome biogenesis, and one of these is known to control core particle composition as well⁷¹. The 19S regulatory particle is composed of six Rpt subunits and 13 Rpn and can be subdivided into two subcomplexes that assemble independently, the base and the lid⁷². Assembly of the regulatory particle base in yeast is orchestrated by at least four distinct assembly chaperones (Nas2, Hsm3, Nas6, and Rpn14). None of them associates detectably with the mature 26S proteasome. These factors are conserved from yeast to human. The regulatory particle base assembles from a set of discrete chaperone-associated base subunit complexes; once assembled, the base binds to the lid and all chaperones are released prior to or during regulatory particle and core particle association. The mechanism of lid assembly and lid-base joining are still poorly understood.

1.2.4.3 Ribosome assembly

In living cells, the ribosome is responsible for the final step of decoding genetic information into proteins. This universal 'translation apparatus' is comprised of two subunits, each of which is a complex assemblage of RNA and proteins. The small 40S subunit (30S in prokaryotes) is responsible for decoding, whereas the large 60S subunit (50S in prokaryotes) carries out polypeptide synthesis. Ribosome biogenesis begins with transcription of pre-rRNA, which undergoes co-transcriptional folding, modification and assembly with ribosomal proteins (r-proteins) to form the two subunits. Assembly of ribosomal subunits in bacteria appears to require few (<25) trans-acting factors (including assembly chaperones). By contrast, assembly of eukaryotic ribosomes is a complicated process that requires the concerted efforts of all three RNA polymerases and >200 trans-acting factors (including assembly chaperones). These aid the assembly, maturation and intracellular transport of ribosomal subunits. The process begins

with the RNA polymerase I transcription of the 35S pre-rRNA which associates with many trans-acting factors and ribosomal proteins to a 90S pre-ribosome. Processing of the pre-rRNA involves a series of endo- and exonuclease events. Cleavage in the spacer region between the 18S and the 5.8S rRNAs leads to the formation of pre-40S and pre-60S particles^{73,74}. RNA polymerase III synthesizes the 5S rRNA, which is incorporated into the pre-60S subunit. After separation of the 90S intermediate into a pre-60S and a pre-40S particle, the two precursors follow primarily independent biogenesis and export pathways. Final maturation of the subunits occurs in the cytoplasm. Several trans-acting factors and export factors associated with pre-60S and pre-40S particles are released before the subunits achieve translation competence. In addition, the final rRNA processing steps occur in the cytoplasm⁷³⁻⁷⁵.

1.2.4.4 Pilus assembly

Bacteria commonly express proteinaceous appendages on their outer surfaces. One class of extracellular polymers, known as pili, is used in attachment and invasion, biofilm formation, cell motility and transport of proteins and DNA across membranes. These non-flagellar appendages of Gram-negative bacteria can be categorized into five major classes on the basis of the biosynthetic pathway involved. Of these five classes, the chaperone-usher pili are the most extensively studied. Chaperone-usher pili are assembled at the outer membrane by two proteins, a periplasmic chaperone and an outer-membrane, pore-forming protein called the usher⁷⁶. The chaperone facilitates folding of pilus subunits, prevents them from polymerizing in the periplasm and targets them to the usher⁷⁷. The usher acts as an assembly platform, recruiting chaperone-subunit complexes from the periplasm, coordinating their assembly into a pilus and secreting that pilus through the usher pore. Pilus subunits are taken up by their cognate periplasmic chaperones as soon as they exit the Sec machinery (which mediates general secretion)⁷⁸. In the absence of the chaperone, the subunits cannot fold properly and form aggregates that are targeted for degradation^{77,79}. The periplasmic chaperones are boomerang-shaped, ~25 kDa proteins consisting of two immunoglobulin-like (Ig-like) domains⁸⁰. Pilus subunits are characterized by an incomplete Ig-like fold that lacks the C-terminal β -strand⁸¹. As a result, all subunits possess a large groove where the missing strand would normally be in a complete Ig fold. In chaperone-subunit complexes, the chaperone inserts a β -strand into the hydrophobic groove of a pilus subunit (donor strand complementation). During subunit polymerization, the complementing β -strand donated by the chaperone is replaced by the N-terminus on the subunit of the incoming chaperone-subunit complex (donor strand exchange). Because of their important role in bacterial virulence, chaperone-usher pili have received considerable attention in vaccine development programmes and in the search for new antibacterials.

1.3 Photosynthesis

Photosynthesis is a fundamental process in biology. By this process, solar energy is being converted into chemical energy. The chemical energy is then used to fix CO_2 and to synthesize the reduced carbon compounds upon which all life depends. Photosynthetic organisms occur in all domains of life (bacteria, archaea, eukaryotes). The compartmentalised process consists of the light-dependent reactions, providing the reducing equivalents and ATP, and the "light-independent (dark) reactions that build up and reduce the carbohydrate chain. In eukaryotes, photosynthesis takes place in specialised organelles called chloroplasts, which are derived from an endosymbiotic event with a cyanobacterium. Most land plants use the C3 pathway of photosynthesis, also called the carbon reduction cycle. C3 plants have a single chloroplast type that performs all of the reactions (Fig.6). This organelle has a special double outer membrane that allows the diffusion of gases as CO_2 and O_2 . The thylakoid membrane, located on the inside of chloroplasts, is highly folded and usually arranged as flattened-stacked vesicles called grana. The machinery necessary for the capture of solar energy and conversion to chemical energy (light-dependent reactions) is embedded in the thylakoid membrane including: photosystems I and II, electron carrier chains, chloroplast ATP-synthetase, light harvesting complexes, as well as other complexes⁸². The aqueous compartment inside the thylakoid membrane is called the lumen, while the aqueous phase outside the thylakoid membrane is the stroma. The stroma contains the soluble proteins and molecules necessary for the reduction of atmospheric carbon into organic carbon (light-independent reactions). Although chloroplasts are thought to originate from a unique primary endosymbiotic event, many algae contain photosynthetically active chloroplasts that are surrounded by three or four membranes. These chloroplasts are called complex plastids and they originated from a secondary endosymbiotic event in which a photosynthetic eukaryotic cell was taken up by a non-photosynthetic eukaryotic host cell⁸².

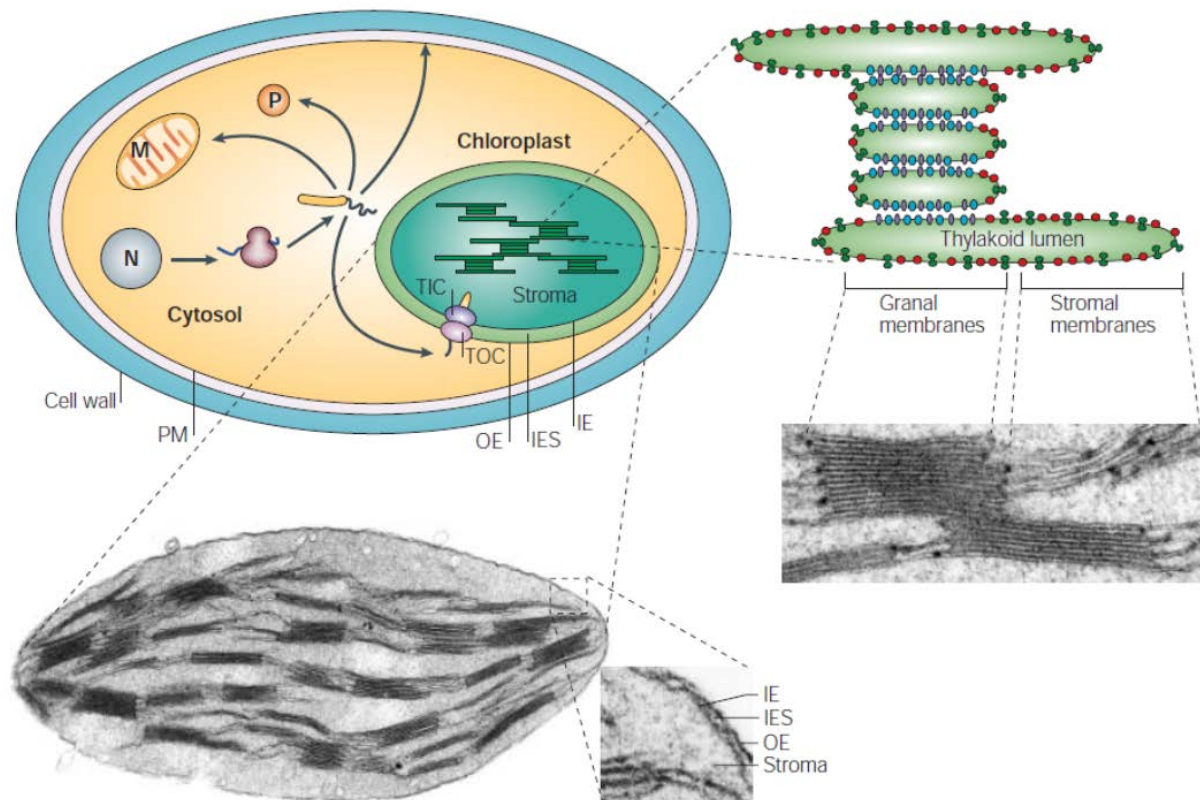


Fig.6: Outline of chloroplast ultrastructure. The chloroplast is composed of three membrane systems: the outer (OE) and inner envelope (IE), and the thylakoid membrane network that contains the protein complexes that are involved in photosynthesis. In addition, three soluble spaces can be distinguished: the inter-envelope space (IES), the stroma and the thylakoid lumen. An electron micrograph of an isolated *Pisum sativum* chloroplast is shown, and the enlargement on the far right shows the typical organization of thylakoids. Granal membranes are preferentially enriched in photosystem II (blue) and the cytochrome b_6f complex (purple), whereas stromal membranes are enriched in photosystem I (red) and the ATP synthase (green). Also shown are the different compartments, a newly synthesized protein can be transported from and to in a cell (N, nucleus; M, mitochondria; P, peroxisomes; PM, plasma membrane; endoplasmatic reticulum not shown). Protein import into chloroplasts is achieved by two translocons called TOC and TIC that reside in the outer and inner envelope, respectively (reproduced from Soll and Schleiff⁸²).

1.3.1 Light-dependent reactions

Photosynthesis can be divided into two processes: light-dependent reactions and light-independent carbon-fixation/assimilation reactions. During the light-dependent reactions, solar energy is being converted into chemical energy in form of NADPH and ATP (Fig.7). Biochemical and biophysical studies showed that the chloroplast thylakoid membrane is capable of light-dependent water oxidation, NADP reduction and ATP formation⁸³. These reactions are catalysed by two separate Photosystems, photosystem I (PSI) and photosystem II (PSII) and an ATP synthase (F-ATPase): the latter produces ATP at the expense of the proton motif force (pmf) that is formed by the light reaction⁸⁴⁻⁸⁸. The cytochrome- $b_6 f$ complex mediates electron transport between PSII and PSI and converts the redox energy into a high-energy intermediate (pmf) for ATP formation⁸⁹. PSI and PSII contain chlorophylls and other pigments that harvest light and funnel its energy to a reaction centre. The reaction center of

photosystem I absorbs light maximally at 700 nm, correspondingly this reactive center is named P700. Whereas, the reaction center of photosystem II absorbs light maximally at 680 nm and is accordingly named P680. Energy that has been captured by the reaction centre induces the excitation of specialized reaction centre chlorophylls, which initiates the translocation of an electron across the membrane through a chain of cofactors. Water, the electron donor for this process, is oxidized to O_2 and 4 protons by PSII. The electrons that have been extracted from water are shuttled through a quinone pool and the cytochrome-*b6 f* complex to plastocyanin, a small, soluble, copper-containing protein⁹⁰. Solar energy that has been absorbed by PSI induces the translocation of an electron from plastocyanin at the inner face of the membrane (thylakoid lumen) to ferredoxin on the opposite side (stroma). The reduced ferredoxin is subsequently used in numerous regulatory cycles and reactions, which include nitrate assimilation, fatty-acid desaturation and NADPH production. The charge separation in PSI and PSII, together with the electron transfer through the cytochrome-*b6 f* complex, leads to the formation of an electrochemical-potential gradient (the pmf), which powers ATP synthesis by the fourth protein complex, F-ATPase⁹¹. In the dark, CO_2 reduction to carbohydrates is fuelled by ATP and NADPH⁹². Cyclic photophosphorylation is an alternative electron-transfer pathway that, unlike the prevailing linear mode, does not involve PSII. In this process, a pmf is formed by electrons that flow from PSI through a quinone and the cytochrome-*b6 f* complex back to PSI. No NADPH is formed in this pathway and its molecular basis and regulation are ill-defined. Cyclic phosphorylation supplies the CBB cycle with extra ATP⁹³.

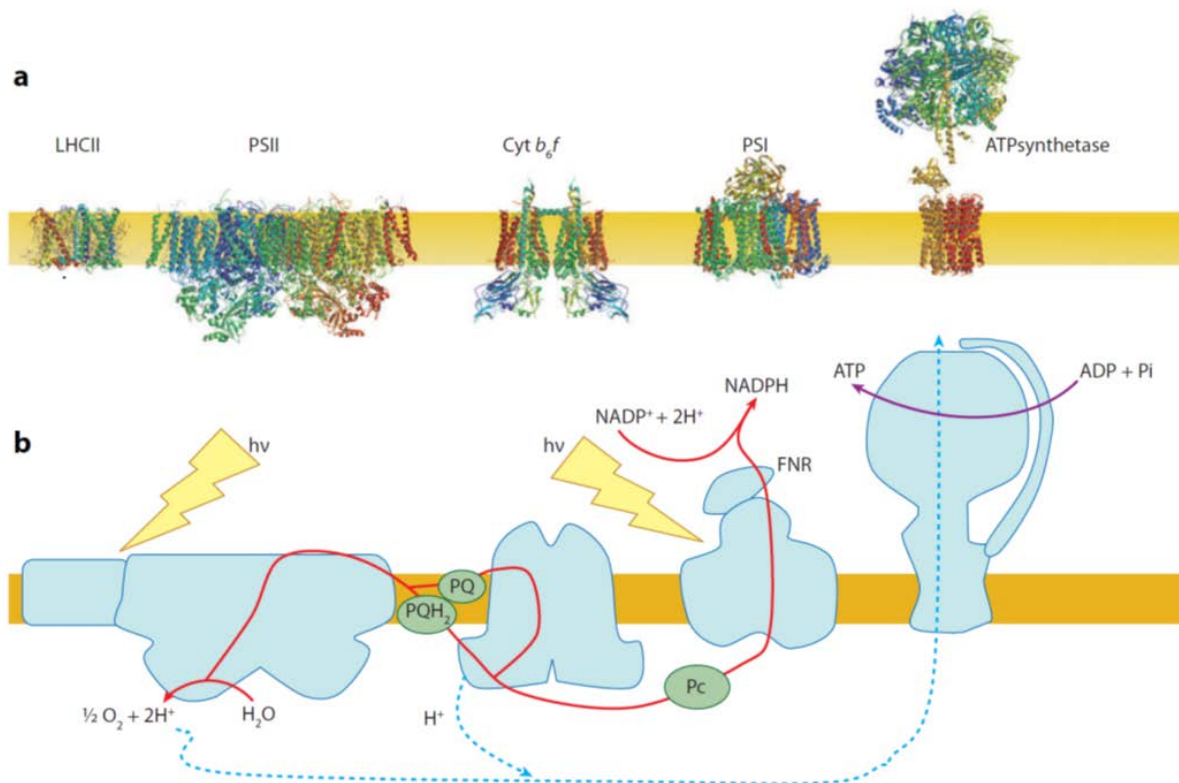


Fig.7: Structures of the large membrane-protein complexes in thylakoid membranes that drive oxygenic photosynthesis. A: Transmembrane organisation of the major photosynthetic proteins in their native oligomerization state. B: Schematic representation of the pathway for photosynthetic linear electron flow and proton translocation through major protein complexes whose atomic structures are shown in a. Electrons are extracted from water on the luminal side of membranes and transferred to NADP on the stromal side of membranes. Electron transfer is driven by the reaction center from two distinct photosystems, PSII and PSI, which are the site of a light-induced charge separation between a photosensitive chlorophyll and an acceptor molecule. The intersystem electron carriers consist of a pool of plastoquinone molecules soluble within the lipid bilayer, a transmembrane protein complex, the cyt b6f complex, comprising an Fe-S cluster and four hemes, a small copper-containing soluble protein in the thylakoid lumen, plastocyanin, which is replaced by a soluble cyt, c6, in some photosynthetic organisms. Protons translocated across the membrane during linear electron flow are used by the transmembrane ATP synthase to drive ATP synthesis (reproduced from Eberhard *et al.*⁹⁴).

1.3.2 Light-independent reactions / Calvin Benson Bassham cycle

The Calvin Benson Bassham cycle (CBB) is the primary pathway of carbon fixation and is located in the chloroplast stroma of higher plants. It comprises 11 different enzymes, catalyzing 13 reactions, and is divided into three phases: carbon fixation, reduction (carbohydrate synthesis), and regeneration of the CO₂ acceptor ribulose-1,5-bisphosphate (RuBP). The carbon fixation reaction is catalyzed by the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) which catalyzes the carboxylation of the CO₂ acceptor molecule RuBP. The 3-phosphoglycerate (3PG) formed by this reaction is then utilized to form the triose phosphates, glyceraldehyde phosphate (G3P) and dihydroxyacetone phosphate (DHAP), via two reactions that consume ATP and NADPH generated in the light reactions. The

regenerative phase of the cycle involves a series of reactions that convert triose phosphates into the CO₂ acceptor molecule RuBP (Fig.8).

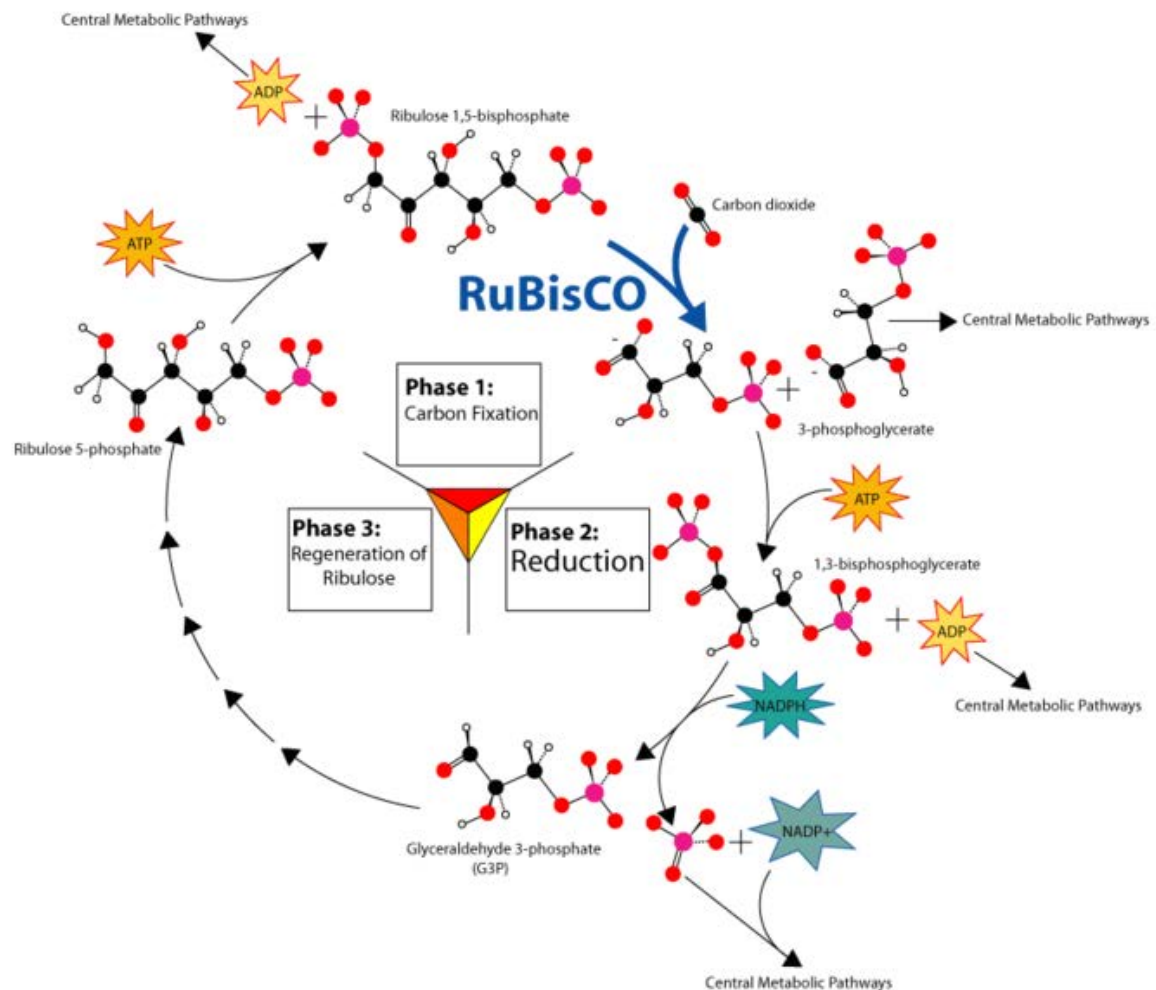


Fig.8: The CBB cycle. The CBB pathway can be divided into 3 different phases. In phase 1, Rubisco catalyzes the addition of CO₂ to the acceptor molecule, RuBP. In Phase 2, the resulting 3PG is reduced to form G3P. This involves the consumption of ATP and NADPH that are created in the light-dependent reactions of photosynthesis. In Phase 3, the regeneration of the acceptor molecule RuBP takes place and involves a series of reactions, which include the enzymes transketolase and transaldolase (not shown here). These reactions require ATP and involve intermediates of sugars with a different number of carbon atoms.

This is accomplished in a series of isomerase, transaldolase and transketolase reactions. In the final step the precursor ribulose 5-phosphate is phosphorylated in an ATP-dependent reaction. Overall the CBB cycle generates one molecule of glyceraldehyde 3-phosphate from 3 molecules of CO₂ at the "metabolic cost" of 9 molecules of ATP and 6 molecules of NADPH. The majority of the triose phosphate produced in the CBB cycle remains within the cycle to regenerate RuBP. However, carbon compounds produced in this cycle are essential for growth and development of the plant and therefore triose phosphates exit from the cycle and are used to synthesize sucrose and starch (Fig.8). The CBB pathway also supplies intermediates to an array of other pathways in the chloroplast, including the shikimate pathway for the biosynthesis

of aromatic amino acids and lignin, isoprenoid biosynthesis and precursors for nucleotide metabolism and cell wall synthesis⁹⁵.

1.3.3 Photorespiration

In the early 1970s it was discovered that Rubisco catalyzes an additional reaction involving molecular oxygen^{96,97}. This enzymatic activity is referred to as ribulose biphosphate oxygenase, and it catalyzes the first reaction in the process of photorespiration, which many plant physiologists regard as the most important metabolic constraint on plant productivity⁹⁸. In this reaction atmospheric O₂ reacts with the acceptor molecule RuBP. Instead of 2 molecules of 3PG (see CBB cycle Fig.8), this reaction results in the formation of one molecule of 3PG and one molecule of 2-phosphoglycolate, a molecule of not only limited use to most organisms but that is also CBB cycle inhibitory and has cellular toxicity character. Conversion of 2-phosphoglycolate into 3PG requires the remainder of the photorespiration pathway, which consumes energy and leads to the release of CO₂ and ammonia (Fig.9). Under today's atmospheric conditions (0.035% CO₂, 21 % O₂, and 78% N₂), the CO₂ concentration in the chloroplasts of C3 plants is approximately 1000 times lower than that of O₂. This unfavorable CO₂/O₂ ratio leads to a significant amount of photorespiration, reducing the overall efficiency of net photosynthesis in C3 plants by approximately one third⁹⁸. Both the carboxylase and the oxygenase reactions of Rubisco occur at the same active site and have the same initial steps, i.e. the binding of RuBP and the generation of an enzyme bound enediol intermediate⁹⁹. CO₂ and O₂ compete to react with this enediol intermediate and once the enediol has reacted with either CO₂ or O₂ the enzyme is committed to form products. ¹⁸O-labeling experiments showed that oxygenation is accompanied by the incorporation of one atom of molecular O₂ into the carboxyl group of 2-phosphoglycolate. The other atom of molecular O₂ is released into the medium⁹⁷. The photorespiratory pathway is comprised of 8 enzymes in four different compartments: chloroplast, cytosol, peroxisome and mitochondrion.

It has been suggested that photorespiration is important for energy dissipation to prevent photoinhibition¹⁰⁰⁻¹⁰³. In addition, photorespiration can generate metabolites, such as serine and glycine, which can be exported out of the leaf¹⁰⁴ or used in other metabolic pathways, for example, provision of glycine for the synthesis of glutathione^{105,106}. Since glutathione is a component of the antioxidative system in plants¹⁰⁷, photorespiration may provide additional protection against oxidative damage in high light by supplying glycine. Thus, photorespiration, despite being energetically wasteful, may also be a useful process in plants¹⁰⁸.

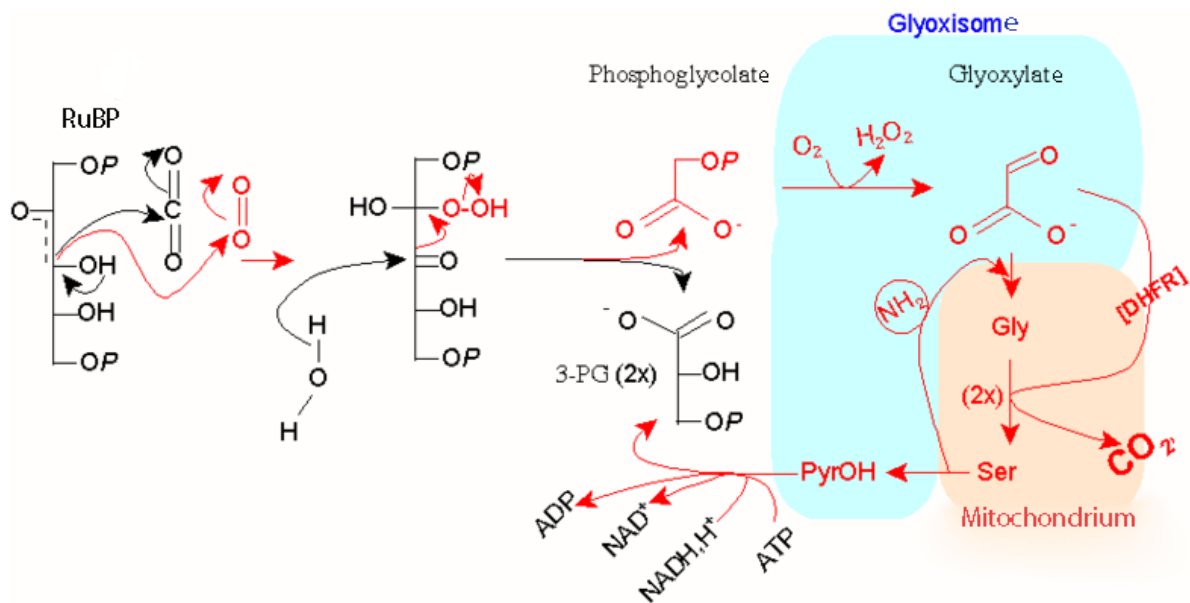


Fig.9: Photorespiratory pathway in plant cells. In photorespiration, RuBP reacts with oxygen to form one molecule of 2-phosphoglycolate that is being hydrolyzed to glycolate. After transport into the peroxysomes (glyoxysome), glycolate is oxidized to glyoxylate. Glyoxylate can be transaminated to glycine. Half of the glycine molecules are converted to N⁵, N¹⁰-methylene tetrahydrofolate (THF) in the mitochondria. In this reaction CO₂ and NH₃ are released. The other half of the glycine molecules can react with N⁵, N¹⁰-methylene THF to form serine. After transport from the mitochondria to the peroxysomes, serine is converted to hydroxypyruvate, which is reduced to glycerate. Glycerate is then phosphorylated and reduced in the chloroplasts to form 3-PG. This reaction consumes ATP and NADPH.

1.3.4 Carbon concentrating mechanisms

Photosynthetic organisms have adapted to the limited availability of carbon dioxide in their environment by developing systems called the carbon dioxide concentrating mechanism (CCM). These include active inorganic carbon (Ci; CO₂ and/or HCO₃⁻) uptake that leads to the accumulation of Ci to levels significantly higher than in the cell exterior. Carbonic anhydrase activity converts much of the accumulated hydrogen carbonate to CO₂, concentrating this substrate around Rubisco and thereby optimising photosynthetic efficiency even under low CO₂ conditions. The efficiency of the process is further improved by the sequestration of Rubisco into specialized structures like the cyanobacterial carboxysome or the pyrenoid in eukaryotic algae. By converting bicarbonate into carbon dioxide in these structures, a local elevation of carbon dioxide concentration is achieved, favoring carboxylation over oxygenation.

1.3.4.1 The Carboxysome

In Cyanobacteria, Rubisco is localized in the carboxysome, a microcompartment enclosed by a proteinaceous, icosahedral shell¹⁰⁹. Carboxysomes can be divided into two types that differ in their component proteins and the organization of their respective genes¹¹⁰. Carboxysomes of the α-type are found in α-cyanobacteria (for example, *Prochlorococcus* species and certain *Synechococcus* species, such as *Synechococcus* sp. WH 8102), whereas carboxysomes of the β-type are found in β-cyanobacteria (for example, *Synechococcus elongatus* PCC7942 and *Synechocystis* sp. PCC 6803). Not all of the genes that are present in one type of

Introduction

carboxysome can be identified in the genomes of organisms that produce the other type, indicating that there could be functional differences between the two types. Organisms with α -carboxysomes arrange their carboxysome genes in a single operon, whereas organisms with β -carboxysome genes are typically arranged in multiple gene clusters. Carboxysomes carry out the final stages of the CCM, through which autotrophic prokaryotes accumulate inorganic carbon to enhance CO_2 fixation (Fig.10). The first part of the CCM involves transmembrane pumps, which actively concentrate bicarbonate inside the cell¹¹¹, whereas the second part involves the carboxysome, in which bicarbonate is converted to CO_2 in the vicinity of Rubisco. Models for how the carboxysome contributes to the CCM predict that a carbonic anhydrase needs to be sequestered with Rubisco in the carboxysome to convert bicarbonate, presumably as it enters the microcompartment from the cytosol, to the Rubisco substrate CO_2 . The outer shell could provide a barrier to the diffusion and leakage of CO_2 and bicarbonate substrates^{112,113}.

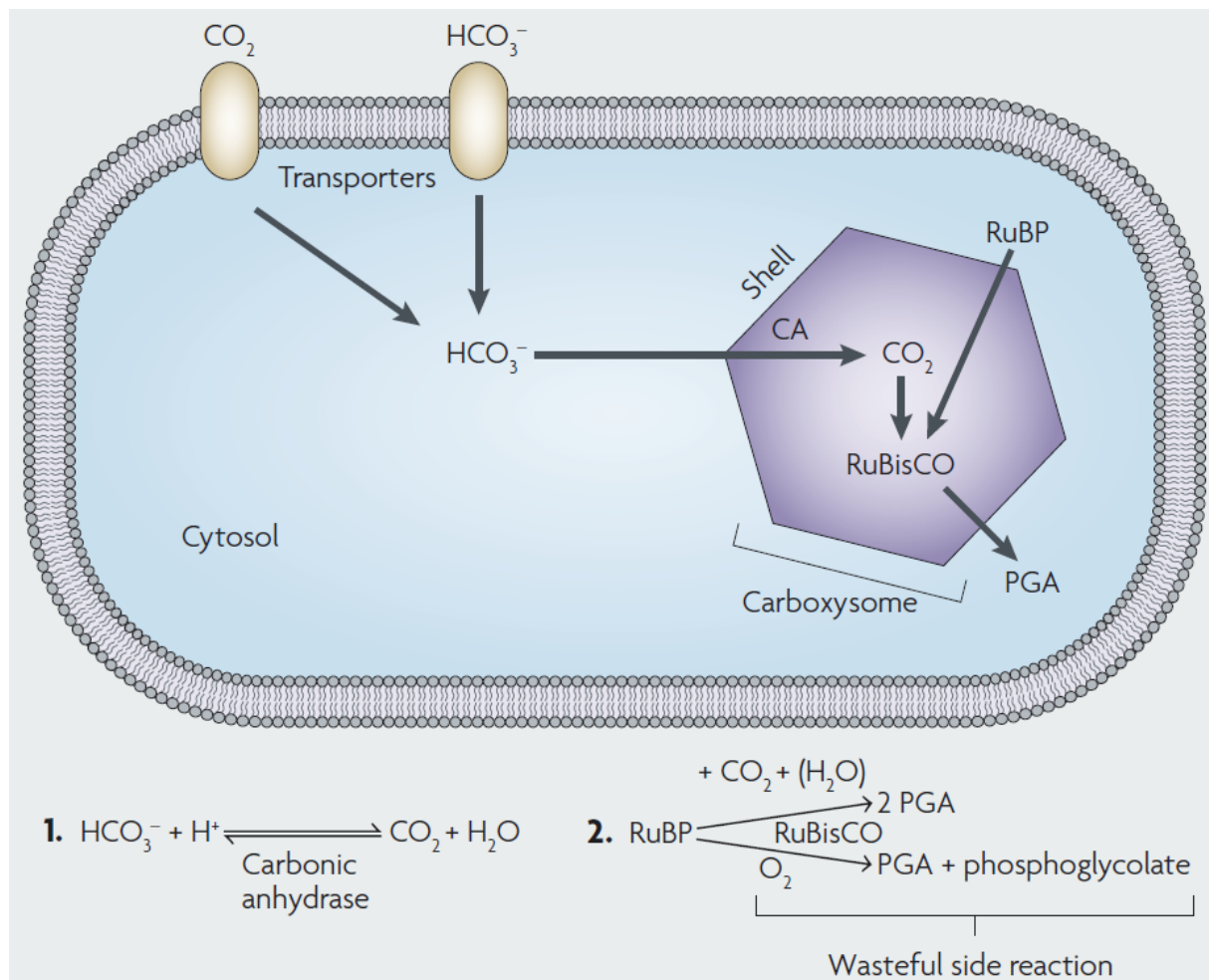


Fig.10: Carbon concentration and the carboxysome. In the first part of the CCM, bicarbonate is concentrated inside the cell by transporters in the cell membrane. The carboxysome is involved in the second part of the CCM, and enhances CO_2 fixation by co-localizing the two enzymes Rubisco and carbonic anhydrase (CA) inside a thin shell that is assembled from thousands of protein subunits. Bicarbonate is thought to enter the carboxysome through pores in the proteinaceous shell, where it is converted to CO_2 for use by Rubisco. RuBP must also enter the carboxysome and the product 3PG must exit the carboxysome (reproduced from Yeates *et al.*¹¹⁴).

1.3.4.2 The pyrenoid

In eukaryotic algae, the pyrenoid serves a function similar to the role of the carboxysome in cyanobacteria. Rubisco is localised to the pyrenoid in chloroplasts of green algae including *Chlamydomonas reinhardtii*¹⁵. Pyrenoids are composed largely of Rubisco and include the formation of a starch sheath around the enzyme instead of a proteinaceous shell. The chloroplast thylakoid carbonic anhydrase plays an essential role in the carbon dioxide concentrating mechanism of *Chlamydomonas*. It catalyzes the conversion of bicarbonate into carbon dioxide and results in a high local concentration of carbon dioxide in the vicinity of Rubisco. Due to the diversity of employed CCM and the difficulty of biochemically characterizing oligosaccharide enclosed microcompartments the mechanistic details of CCM in algae are mostly unknown.

1.3.4.3 C4 and crassulacean acid metabolism (CAM) photosynthesis

C4 and CAM photosynthesis are evolutionarily derived from C3 photosynthesis. The morphological and biochemical modifications necessary to achieve either C4 or CAM photosynthesis are thought to have independently arisen numerous times within different higher plant taxa. It is thought that C4 photosynthesis evolved in response to low atmospheric CO₂ concentrations, which significantly increases photorespiration rates in C3 plants, thus reducing photosynthetic productivity; both C3-C4 intermediate and C4 plants exhibit reduced photorespiration rates. In contrast, it may be argued that CAM arose either in response to selection of increased water-use efficiency or for increased carbon gain. Globally, all three pathways are widely distributed today, with a tendency toward ecological adaptation of C4 plants into warm, monsoonal climates and CAM plants into water-limited habitats.

In C4 plants, a simple change in expression of the C3 cycle is used to overcome the reduced photosynthetic efficiency associated with photorespiration. Instead of allowing the C3 photosynthesis cycle to occur in all photosynthetic cells, the C3 cycle is limited to selected interior cells, typically the bundle sheath cells. A layer of mesophyll cells surrounds these bundle sheath cells. Within the mesophyll cells is phosphoenolpyruvate (PEP) carboxylase, an enzyme that catalyzes the initial photosynthetic reaction. This reaction involves phosphoenolpyruvate and atmospheric CO₂ as substrates to produce oxaloacetate, a four-carbon acid; hence the name C4 photosynthesis. The C4 acid diffuses from the mesophyll through plasmodesmata to the bundle sheath cells, where the C4 acid is decarboxylated. Since PEP carboxylase has a higher affinity for its substrate and a greater maximum velocity than Rubisco, the CO₂ concentration in the bundle sheath cells ends up being significantly higher than that in either the mesophyll cells or the surrounding atmosphere. The consequence is that PEP carboxylase effectively serves as a CO₂ pump, concentrating CO₂ within the bundle sheath cells. This results in CO₂ concentrations within the chloroplasts that are an order of

magnitude higher than in C3 plants. As a consequence, the Rubisco reactions in C4 plants take place in an atmosphere with a high CO₂/O₂ ratio, and photorespiration is effectively eliminated. Critical to the functioning of the C4 cycle is a distinct spatial separation of the activities of Rubisco and PEP carboxylase (Kranz anatomy)-such that with PEP carboxylase activity located between atmospheric CO₂ and Rubisco, it provides a pump to raise the CO₂ concentrations internally¹¹⁶.

CAM photosynthesis involves the same CO₂ concentrating mechanism as in C4 photosynthesis (PEP carboxylase). However, rather than a spatial separation of the two carboxylation enzymes such as exists in C4 photosynthesis, there is a temporal separation of the two carboxylases in CAM, and both reactions occur within the same cell. In contrast to C3 and C4 plants, an inverted diurnal pattern of stomatal opening occurs in CAM plants. Stomata open during the evening, and atmospheric CO₂ is fixed as a C4 acid (malate), which is then stored within the vacuole of the photosynthetic cell. During the following day, stomata remain closed, and malate is decarboxylated to release CO₂. The CO₂ concentration within the cell remains high as CO₂ cannot diffuse out through the closed stomata. As a result, Rubisco once again operates within a high CO₂/O₂ environment, and photorespiration is eliminated. Thus, in CAM plants a temporal separation of the two carboxylase reactions occurs with Rubisco activity within the cell during the day and PEP carboxylase activity during the night¹¹⁷.

1.4 Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco)

Much of the current diverse interest in D-ribulose-1,5-bisphosphate carboxylase/oxygenase can be attributed to its pivotal function in photosynthesis and photorespiration. It is clearly a rate-limiting factor in both processes as well as it is an inefficient catalyst. This is very surprising since this enzyme must have been subject to the most intense kind of selection for catalytic specificity and turnover rate¹¹⁸. This inefficiency is the reason that photosynthetic cells must invest one quarter or more of their precious nitrogen resources in this one enzyme, which is the world's most abundant protein; it occurs in all photosynthetic leaves and makes up to 65% of the total soluble protein in leaf extracts. In the stromal compartment of chloroplasts it reaches a concentration of up to about 300 mg/ml¹¹⁹. Rubisco connects the inorganic and organic phases of the biosphere's carbon cycle, catalyzing the only quantitatively significant reaction by which atmospheric CO₂ may be acquired by living organisms. This distinguishes Rubisco from the many other carboxylases. The reactions these enzymes catalyze do not result in the net acquisition of carbon for gluconeogenic purposes or are quantitatively unimportant to the biosphere¹²⁰⁻¹²².

1.4.1 Structure of Rubisco

Due to the possibility of purifying relatively large amounts of the enzyme from plant material, Rubisco was an early candidate for structural studies by X-ray crystallography with the first

25

crystallization reported in 1971¹²³. Nevertheless, the road to a successful structure determination has been littered with pitfalls linked to the fact that the enzyme is unusually large and complicated. That is a reason why it took another one and a half decades until the first X-ray structure of Rubisco was determined to 2.9 Å resolution¹²⁴ from the recombinant dimeric enzyme from *Rhodospirillum rubrum*¹²⁵. The structures from spinach^{126,127}, tobacco^{128,129} and *Synechococcus* followed suit^{130,131}. Rubisco from higher plants, algae and cyanobacteria is a hexadecamer of molecular mass 550 kDa composed of eight large subunits (RbcL: 50-55 kDa) and eight small subunits (RbcS: 12-18 kDa). This type of Rubisco structure is called form I Rubisco. The molecule exhibits 422 symmetry and consists of four RbcL₂ dimers arranged around a four-fold axis, capped at both ends by four small subunits. Based on amino acid sequences of the form I enzyme, a distinction has been made between green-type enzymes (forms I A and B from cyanobacteria, eukaryotic algae and higher plants) and red-type enzymes (forms I C and D from non-green algae and phototrophic bacteria)^{132,133}. Rubiscos from some dinoflagellates and purple nonsulfur bacteria (e.g. *Rhodospirillum rubrum*) are homodimers of two such RbcL subunits related by a twofold rotational symmetry, lacking small subunits. They are called form II Rubiscos. More recently the structure of Rubisco from the thermophilic archaea *Thermococcus kodakaraensis* has been solved¹³⁴, revealing a completely new pentameric quaternary structure, which has been assigned as form III Rubisco. However, in terms of quaternary structure the archaea are diverse and comprise RbcL₂, RbcL₈ and RbcL₁₀ enzymes. The newest member of the Rubisco super family, form IV Rubisco-like-protein (RLP), is also dimeric and solely composed of large subunits. RLP occur in organisms that do not have a CBB cycle. This class is divided into six clades based on primary sequences¹³⁵. The enzyme is termed RLP since it does not harbor any carboxylase activity due to critical substitutions in the active site¹³⁶. The function of RLP is not known for all organisms in which it is found, but it is involved in sulfur metabolism¹³⁷. In *Geobacillus kaustophilus* and *Bacillus subtilis* the RLP is implicated in the methionine salvage pathway and functions as an enolase, stabilizing the enol form of active site substrate¹³⁸. Figure 11 displays the quaternary structures of forms I-III of Rubisco. Despite apparent differences in amino acid sequence between the various forms of Rubisco, the secondary structure is extremely well conserved throughout.

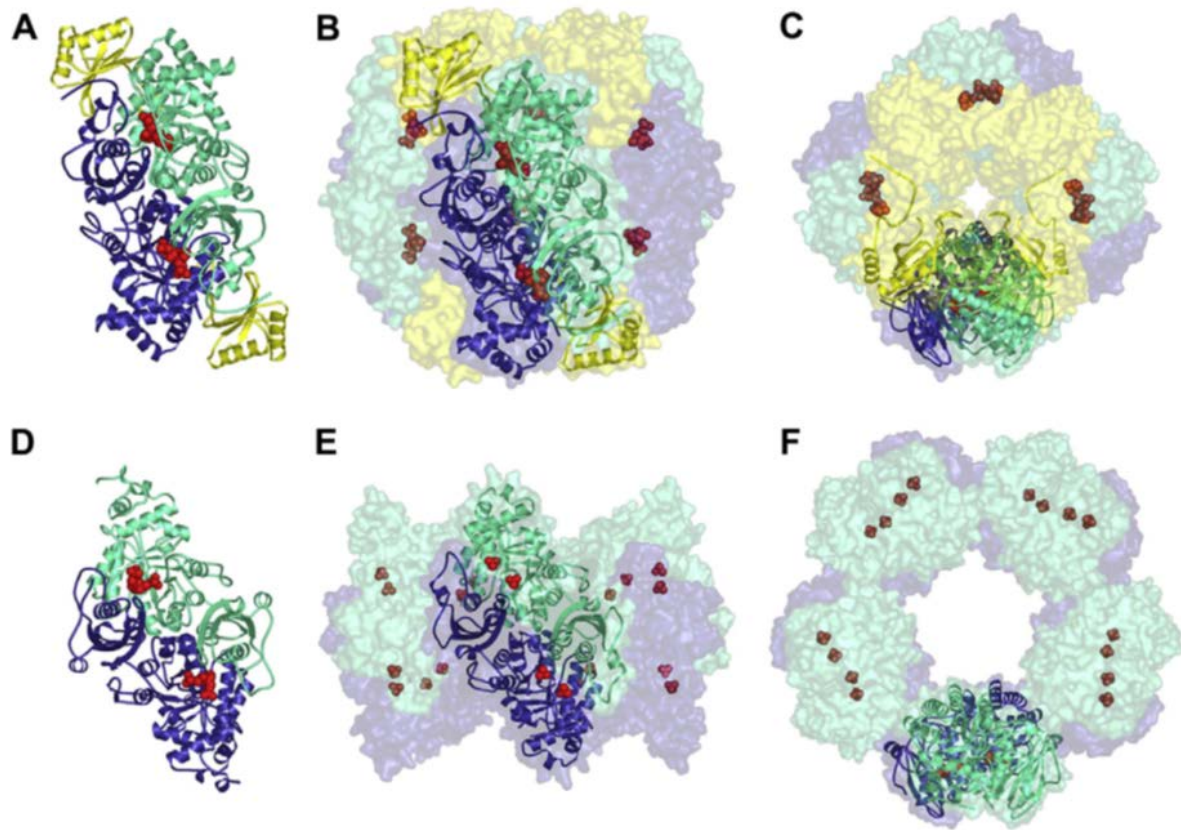


Fig.11: The different quaternary structures of Rubisco isoforms. A: The RbcL₂RbcS₂ unit of form I Rubisco from spinach viewed along the 2-fold symmetry axis. Large subunits are blue and green, small subunits are yellow, and the transition-state analog (CABP) is displayed as red spheres. B: The entire RbcL₈RbcS₈ hexadecamer viewed along the same 2-fold axis and C: along the 4-fold axis. D: The dimeric form II Rubisco from *Rhodospirillum rubrum* showing the 2-fold symmetry. E: and F: The RbcL₁₀ Rubisco from *Thermococcus kodakaraensis* viewed along the 2-fold and 5-fold axes, respectively. Sulphate ions bound in the active site are displayed as red spheres (reproduced from Andersson and Backlund¹⁴³).

The overall fold of the large (catalytic) subunit is similar in all forms of Rubisco (Fig.11): a smaller amino-terminal domain consisting of a four-to-five-stranded mixed β sheet with helices on one side of the sheet and a larger carboxyterminal domain¹³⁹. The carboxyterminal domain consists of eight consecutive $\beta\alpha$ -units arranged as an eight-stranded parallel α/β barrel structure. The active site is located at the carboxyterminal end of the β -strands, with the loops connecting the $\beta\alpha$ -units contributing several residues involved in catalysis and substrate binding. Residues from the aminoterminal domain of the adjacent large subunit in the dimer complete the active site. Thus, the functional unit structure of Rubisco is an RbcL₂ dimer of large subunits harboring two active sites (Fig.11). The secondary structure topology of the large subunit is well conserved (Fig.12).

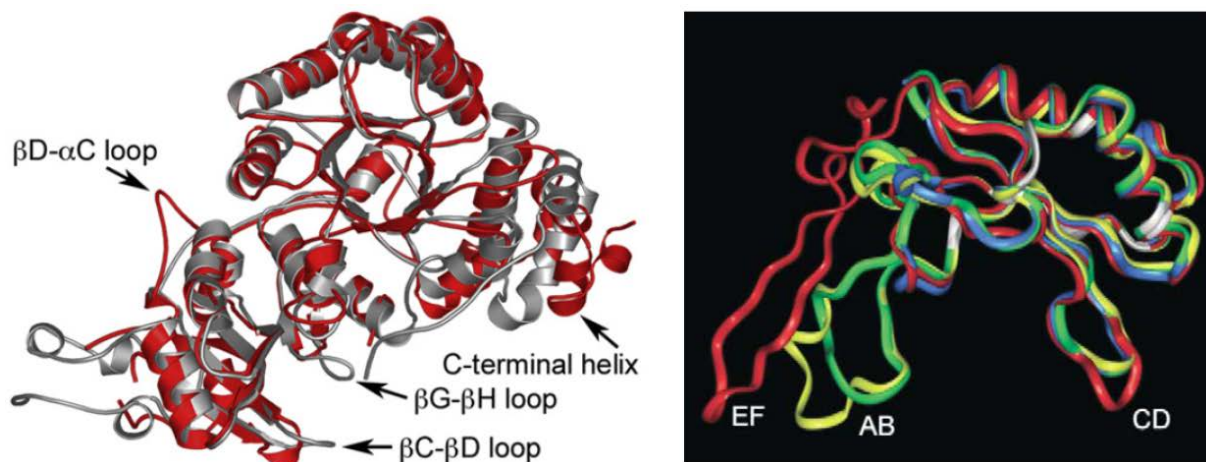


Fig.12: Left: Comparison of the large subunit from spinach (PDB 8RUC, grey) with the large subunit of *Rhodospirillum rubrum* (PDB 5RUB, red). Features that differ are indicated in arrows. Helices 7 and 8 of the α/β barrel are displaced in the enzyme from *Rhodospirillum rubrum* (reproduced from Andersson and Backlund¹⁴³). **Right: Comparison of the Rubisco small-subunit X-ray crystal structures from *Chlamydomonas* (yellow), *Spinacia* (green), *Synechococcus* (blue) and *Galdieria partita* (red).** Loops are labeled relative to their flanking β strands. Residues that are more than 95% conserved among all known small-subunit sequences are colored white (reproduced from Spreitzer¹⁴¹).

Form II Rubisco from *Rhodospirillum rubrum* has an extra α -helix in the amino-terminal domain (helix α A, residues 14-19)¹⁴⁰. In addition, certain loop regions differ in length (Fig.12)¹⁴⁰. Whereas the large subunits display relatively little variation in the different forms, the small subunits are more diverse. The common core structure consists of a four-stranded anti-parallel β -sheet covered on one side by two helices (Fig.12)¹²⁷. The most striking variations occur in two distinct locations, the loop between β strands A and B of the small subunit, the so-called β A- β B-loop, and the carboxyterminus¹⁴¹. The β A- β B loops of four small subunits line the openings of the solvent channel that traverses the holoenzyme along the 4-fold axis (Fig.12). Rubisco from prokaryotes and non-green algae have a loop length of only ten residues (Fig.13A) as illustrated by the structure of the cyanobacterial enzyme¹³⁰, but Rubisco from higher plants have 22 (Fig.13B)¹⁰⁷ and green algal Rubisco have 28 (Fig.13C)¹⁴². Judging by the arrangement of the small subunits covering a substantial area at two opposite ends of the RbcL-subunit octamer, it is reasonable to assume a structural function of the small subunit, namely to assemble and concentrate the large catalytic subunits. However, considering that some Rubisco enzymes lack small subunits and have the lowest specificity values¹⁴⁴, it is tempting to speculate that the small subunits also contribute substantially to the differences in kinetic properties observed among different Rubisco enzymes, despite being spatially separated from the active sites.

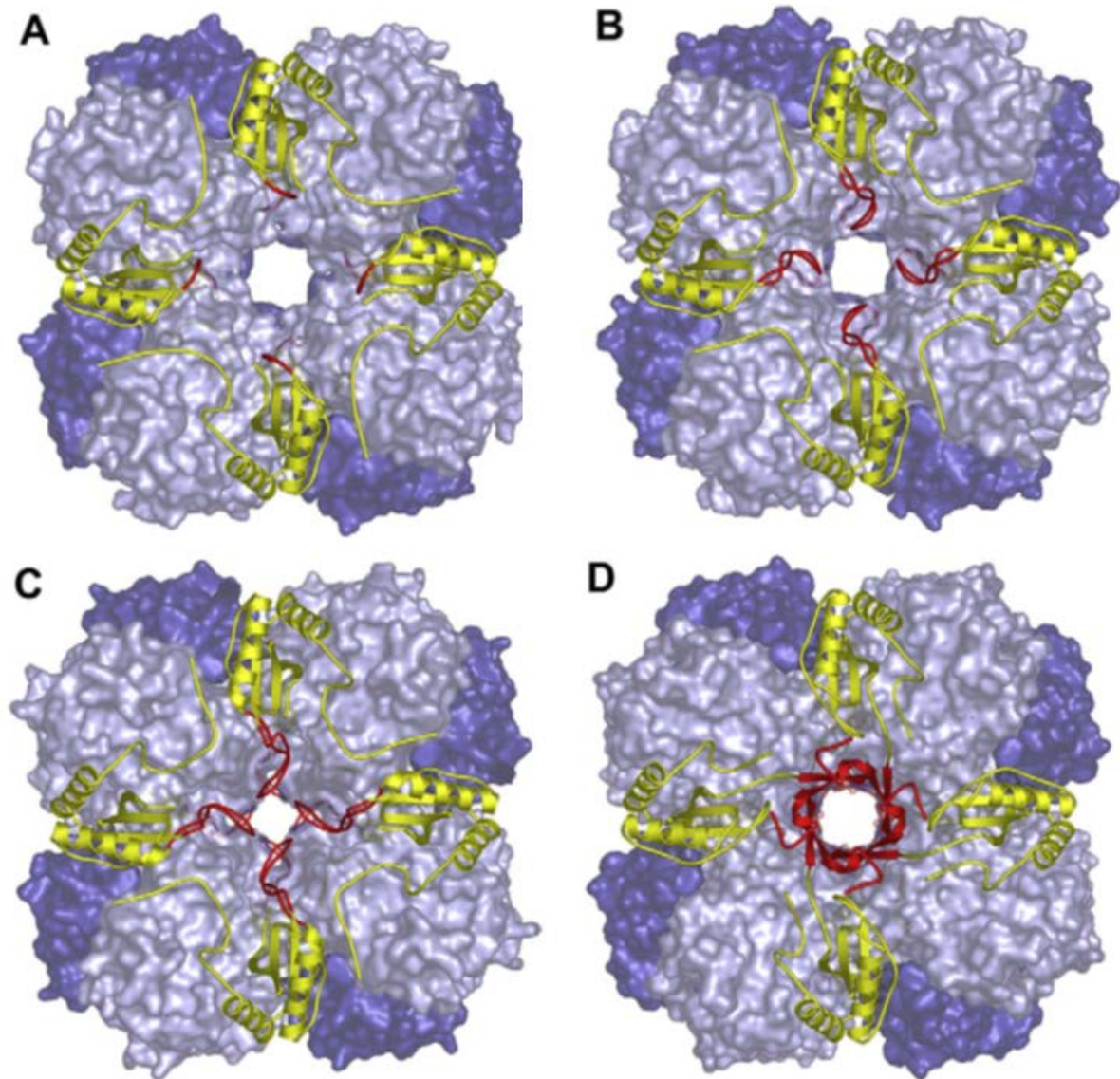


Fig.13: Variation of the β A- β B loop of the small subunit in various Rubisco enzymes. Large subunits are coloured blue, small subunits are yellow, and the β A- β B loops are red. A: *Synechococcus* sp. PCC 6301 (PDB 1RBL). B: *Spinacia* (PDB 8RUC). C: *Chlamydomonas reinhardtii* (PDB 1GK8). D: *Galdieria partita* (PDB 1BWV, reproduced from Anderson and Backlund¹⁴³).

1.4.2 Catalytic mechanism and regulation of Rubisco activity

The catalytically competent form of all Rubiscos, irrespective of their origin, is a ternary complex of enzyme- CO_2 - Mg^{2+} ^{145,146}. The formation of the active ternary complex involves the slow, reversible reaction of a molecule of CO_2 with the ϵ -amino group of Lysine 201 (K201) to form a carbamate (Fig.14)^{147,148}. The activator CO_2 is different from the molecule of substrate CO_2 ^{149,150}.

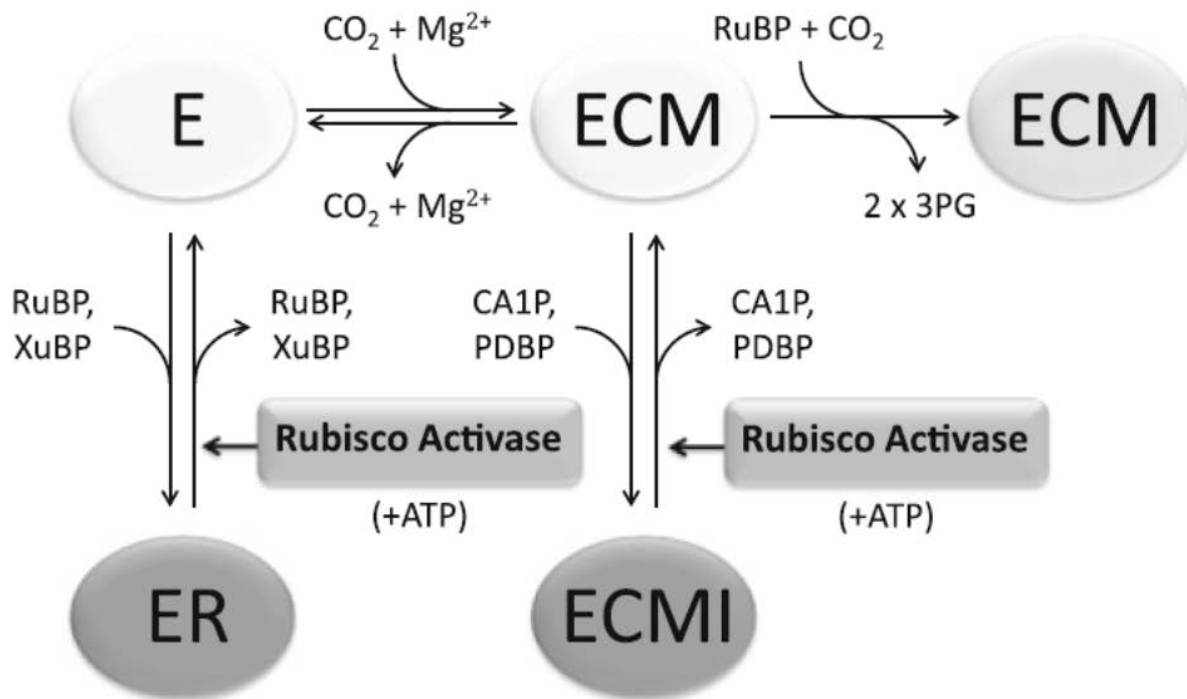


Fig.14: Regulation of Rubisco activity by reversible carbamylation and inhibitor binding to carbamylated or non-carbamylated Rubisco. Rubisco activase and small molecule inhibitors control the activity of Rubisco. E and ECM signify non-activated and active Rubisco, respectively. ER and ECMI are inhibited forms (reproduced from Müller-Cajar *et al.*¹⁵¹).

Formation of the carbamate is followed by the rapid addition of Mg^{2+} to create the active ternary complex. This complex does not require the presence of small subunits¹⁵². In place of Mg^{2+} , Rubisco can accommodate Mn^{2+} , Fe^{2+} , Ca^{2+} , or Cu^{2+} as activating metal ions^{153,154}. The $\text{RbcL}_8\text{RbcS}_8$ Rubiscos contain eight catalytic sites. Both RuBP and various other effectors interact with the enzyme at a common site, the catalytic site. This results in the fact, that all of the compounds that influence the carbamylation state of the enzyme are competitive inhibitors of catalysis with respect to RuBP¹⁵⁵. Hatch and Jensen¹⁵⁶ classified effectors into two groups. Positive effectors, such as NADPH and 6-phosphogluconate, enhance carbamylation. Negative effectors, ribose-5-phosphate for example, favor the decarbamylated state. During Rubisco's multistep catalytic reaction, protonation and oxygenation of the RuBP enediolate intermediate can result in the formation of isomeric pentulose biphosphates, so-called misfire by-products¹⁵⁷. These include xylulose-1,5-bisphosphate (XuBP), 2,3-pentodiulose-1,5-bisphosphate (PDBP) and 3-ketoarabinitol-1,5-bisphosphate (KABP)^{157,158}, whose formation results in an inactive, 'closed' enzyme that reactivates only slowly, limited by the spontaneous opening of the active site (Fig.14). The latter process is accelerated by the activity of another protein, Rubisco activase (Rca), a member of the AAA+ protein family, by an ATP-dependent mechanism (Fig.14)^{159,160}. Interestingly, the activity of Rubisco activase is controlled by redox-regulation in many plant species, which allows for indirect control of Rubisco activity in response to the redox status of the chloroplast stroma due to illumination. In many species in darkness and low light, carboxyarabinitol 1-phosphate (CA1P) is responsible for the low activity

of Rubisco by binding to its activated form¹⁶¹. Presumably in the plant, as photosynthesis slows in shade or dark conditions, this intermediate analog stabilizes the activated state of the enzyme. CA1P is released by Rubisco activase¹⁶² after which it is rendered non-inhibitory by a specific, redox-modulated phosphatase^{163,164}. In the case of XuBP, a specific hydrolase, CbbY, cooperates with the action of Rca by degrading the released inhibitor, thereby preventing its rebinding¹⁶⁵. In the inactivated state, the enzyme can be also inhibited by its natural substrate RuBP, which can accumulate to high levels. One positive consequence of these naturally occurring stable binary states might be the protection of flexible elements from proteolysis during periods of low photosynthetic activity. Once Rubisco is activated, the stage is set for catalysis. The reaction mechanism is ordered, with RuBP binding before addition of the gaseous substrates CO₂ or O₂¹⁶⁶. The catalytic process of carboxylation and oxygenation involve a sequence of analogous intermediates except for a final protonation that is lacking in the case of oxygenation (Fig.15).

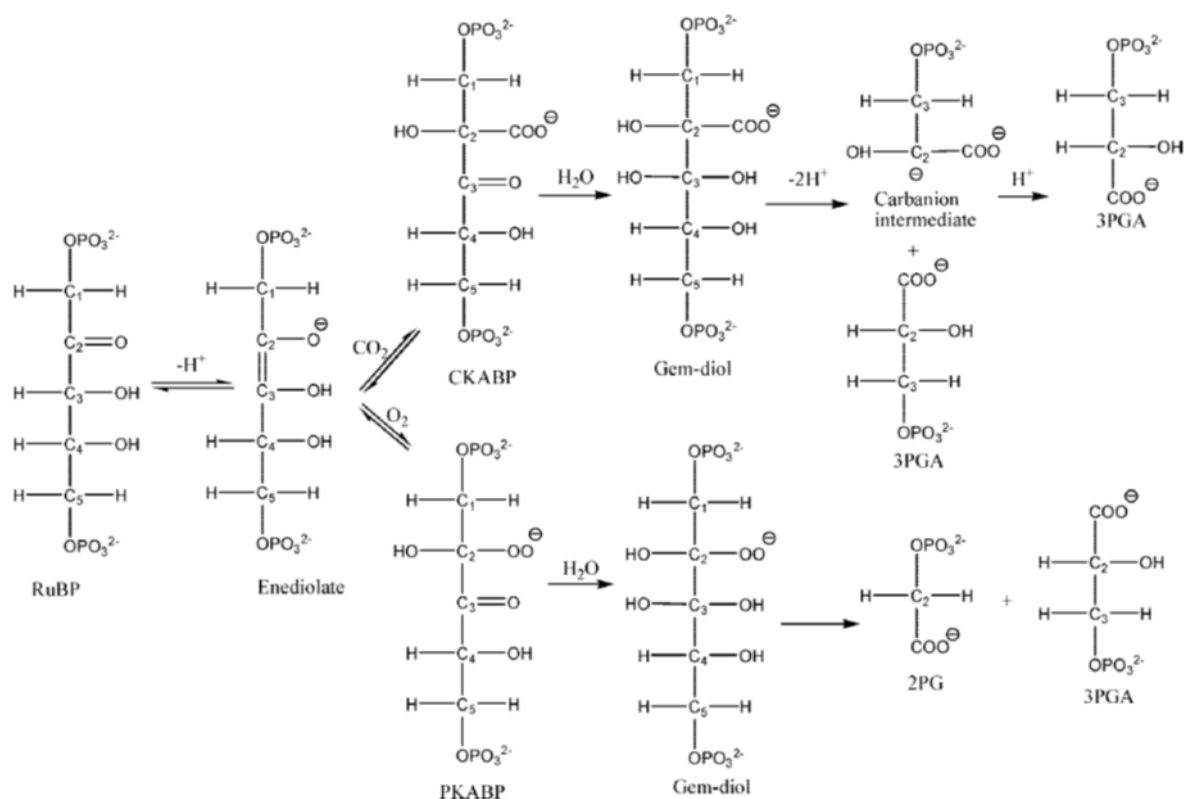


Fig.15: Sequence of reactions catalyzed by Rubisco. RuBP: D-ribulose 1,5-bisphosphate, CKABP: 2-carboxy-3-keto-D-arabinitol-1,5-bisphosphate, PKABP: 2-peroxo-3-keto-D-arabinitol-1,5-bisphosphate, 3PGA: 3-phospho-D-glycerate, 2PG: 2-phosphoglycolate (reproduced from Kannappan and Gready¹⁶⁷).

The initial enolization of RuBP is common to both pathways. CO₂ or O₂ then compete for the resulting enediol producing either a carboxyketone (CKABP, Fig.15) or a peroxyketone (PKABP, Fig.15), respectively. These ketones are hydrated, either in concert with the addition of the gases or subsequently. The hydrated ketones (gem-diol in Fig.15) then split

heterolytically between C-2 and C-3. In the case of oxygenation, this completes the reaction. Carboxylation involves one further step: a proton must be added to the *Si* face of C-2 of the *aci*-acid (carbanion intermediate in Fig.15) produced from C-1 and C-2 of RuBP and the incoming CO₂ molecule in order to produce the second molecule of 3PGA. The active site of Rubisco catalyzes all five reactions, i.e. enolization, gas molecule addition, hydration, C-C cleavage and protonation, without reopening at intermediate steps. The active site is formed from elements of the C-terminal barrel domain of one RbcL-subunit and the N-terminal domain of the second RbcL-subunit in the dimer. In the inactive enzyme, the site is open and accessible to activating cofactors and bisphosphate substrate¹²⁸. After formation of the essential carbamate and coordination of the Mg²⁺, RuBP binds and a series of loops close over the site to enfold and capture the bisphosphate^{127,131}. Closure of the loops brings together amino acids that are critical for catalysis and determine the fate of the substrate (Fig.16). The region of the barrel that comprises the active site is at the C-terminal end of the eight β -strands that form the core of this domain. The loops that connect the strand and helical elements of the barrel extend above and over the surface of the domain, contributing the amino acids that form the 1- and 5-phosphate binding sites and an extensive hydrogen bonding network with the sugar backbone of RuBP.

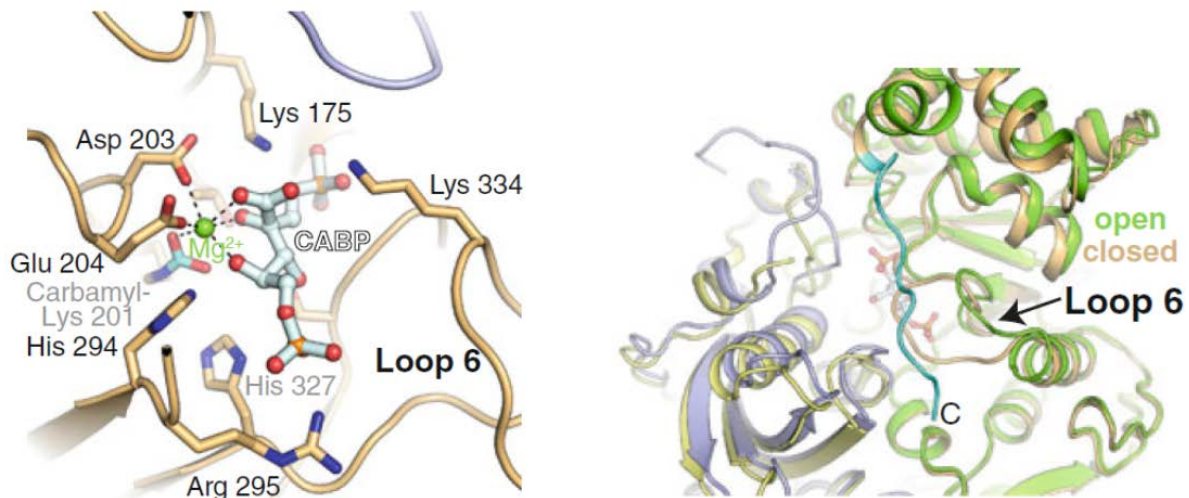


Fig.16: Cross-sectional view through the active site of tobacco Rubisco. Left: residues that contact the transition-state analogue CABP (silver) are shown in stick representation. Oxygen, nitrogen and phosphorus atoms are coloured red, blue and orange, respectively. The catalytic Mg²⁺ ion is shown as a green sphere. One of its ligands is the carbamylated lysine residue 201 (cyan). The active site is at the interface between two large Rubisco subunits (coloured brown and blue). Loop 6 (on the right) seals the binding pocket from the solvent. Right: open and closed conformations of Rubisco. The superposition of the crystal structures for the open (PDB 1EJ7) and closed (PDB 4RUB) states of Rubisco reveals conformational changes around the active site. In the closed conformation (brown and blue ribbons), the C-terminal peptide highlighted in cyan locks down loop 6 (from the right) on top of the active site indicated by CABP (stick representation), whereas in the open conformation (green and yellow ribbons) the C-terminal peptide is oriented away from the active site in a flexible conformation and loop 6 is detached (reproduced from Müller-Cajar *et al.*¹⁵¹).

As mentioned above, in C3 plants, a number of serious limitations to the efficiency of photosynthesis are caused by the catalytic properties and regulation of Rubisco. There are two major factors affecting the carboxylase vs. oxygenase reaction catalyzed by Rubisco: the relative concentration of CO₂ and O₂ at the active site and the ability of the enzyme to discriminate between the two gaseous substrates. The latter one is expressed by the specificity factor $S_{C/O} = (k_{cat}^C/K_C)/(k_{cat}^O/K_O)$, where k_{cat}^C , k_{cat}^O are the maximum velocities and K_C , K_O the Michaelis constants for CO₂ and O₂, respectively¹⁴⁵. Rubisco has very low substrate turnover rates of only ~3-10 CO₂ molecules per second. The specificity factors vary among different Rubiscos, the lowest for *Rhodospirillum rubrum* (12). Moderate values are found in culture plants (e.g. *Nicotiana tabacum* with 82), but the highest are found in marine red algae (e.g. *Griffithsia monilis* with 167). Calculations of a “perfect” Rubisco with an optimal catalysis rate and fixing exclusively CO₂, suggest that this enzyme could maintain photosynthesis with 86% less water loss, 35% less light, and 99% less protein investment in Rubisco¹⁶⁸. Therefore, one strategy to increase photosynthetic carbon fixation would be to generate C3 crop plants expressing Rubisco with high specificity factor and catalytic rate¹⁶⁹⁻¹⁷¹. Attempts to introduce *rbcL* and *rbcS* operons for the high specificity factor red-type Rubiscos of *Galdieria sulphuraria* and *Phaeodactylum tricornutum* into the plastid genome of tobacco have only partially succeeded¹⁷². Whilst the transgenes directed the synthesis of transcripts in abundance, the subunits of these foreign Rubiscos were insoluble, indicating that one or more processes associated with the folding and/or assembly of the red-type Rubisco will need to be understood much better and perhaps translated into chloroplasts if this type of Rubisco transplantation is to be successful.

1.4.3 Rubisco: synthesis, folding and assembly

In green algae and higher plants, Rubisco occurs in the chloroplast and the RbcL subunit is encoded in the chloroplast genome. A family of nuclear genes encodes the RbcS subunit, which is synthesized as a precursor polypeptide on cytosolic ribosomes and imported into the chloroplast in an ATP-dependent reaction¹⁷³. The N-terminal signal peptide of the RbcS subunit is proteolyzed prior to assembly with the RbcL subunits^{174,175}. The RbcL gene in higher plants is present as a single copy per chloroplast genome, but because many copies of the genome are present in each plastid, the actual RbcL copy number per chloroplast can be high. Assembly of both RbcL₂ and RbcL₈RbcS₈ Rubisco from photosynthetic prokaryotes and rhodophytic algae is apparently simpler, given that the genes encoding their subunits are found within the same operon and that synthesis and assembly are cytosolic/stromal events^{176,177}. After import into chloroplasts and processing, the mature RbcS subunits are assembled with plastid synthesized RbcL subunits into the RbcL₈RbcS₈ holoenzyme^{174,178}. Before assembly, they might be associated with cpn60^{179,180}. Although Rubisco RbcL subunits synthesized in

isolated chloroplasts can assemble into holoenzyme, a significant proportion of RbcL subunits are also stably associated with cpn60¹⁸³. As newly synthesized RbcL subunits assemble into Rubisco, the pool of RbcL subunits bound to cpn60 declines. This raises the possibility that nascent Rubisco RbcL subunits are specifically associated with cpn60 before assembly of Rubisco^{180,181}. In addition to interacting with Rubisco subunits, chloroplast cpn60 can bind to many different imported proteins, and it clearly plays a general role in chloroplast biogenesis^{182,183}. There have been no reports of imported or stromal synthesized Rubisco subunits binding to the chloroplast heat shock protein 70 chaperone *in vivo*, but this is clearly a possibility, given that other imported proteins can interact with heat shock protein 70^{184,185}. Furthermore, it has been reported that the 70 kDa heat shock protein/DnaK chaperone system is required for the productive folding of cyanobacterial Rubisco subunits in *E. coli*¹⁸⁶. As shown in cyanobacteria, assembly of the RbcL₈RbcS₈ complex is thought to involve the formation of RbcL₈ core particles, followed by the docking of unassembled small subunits¹⁸⁷. Experiments designed to examine the assembly of Rubisco in isolated chloroplasts show that an initial binding step between cpn60 and newly synthesized or imported Rubisco subunits is detectable^{179-181,188,189}. Like chloroplast cpn60, the GroEL oligomer binds to newly synthesized RbcL subunits. When prokaryotic RbcL₂ and RbcL₈RbcS₈ forms of Rubisco are synthesized in *E. coli*, successful assembly requires GroEL and GroES proteins¹⁹⁰. To understand the mechanism of action of the GroE proteins on Rubisco folding, an *in vitro* refolding assay was developed that consists of purified GroEL and GroES proteins, Mg²⁺ATP, K⁺ ions, and chemically denatured and unfolded *Rhodospirillum rubrum* form II Rubisco¹⁹¹. Spontaneous chaperonin-independent reconstitution of Rubisco at lower temperatures is inhibited by GroEL binding, which leads to the formation of a stable binary complex¹⁹². Discharge of the GroEL-RbcL complex, which then results in active Rubisco dimers, requires the GroES cochaperonin¹⁹¹. Chloroplast cpn21 effectively substitutes for GroES in the chaperonin facilitated refolding of denatured *Rhodospirillum rubrum* Rubisco, indicating functional similarity¹⁹³. Higher plant Rubisco subunits expressed in *E. coli* do not form active enzyme¹⁹⁴. Although large subunits from higher plants bind to GroEL, they fail to assemble properly. Even when the genes for higher plant Rubisco are co-expressed in *E. coli* with the chloroplast cpn60 genes, no assembly of Rubisco occurs^{195,196}. An attempt to address this issue with anti-sense RNA in tobacco plants showed that the synthesis of a β subunit of cpn60 could be inhibited and the plants showed a variety of defects, which nonetheless did not prevent Rubisco activity¹⁸³. However, Rubisco protein levels and/or assembly states were not analyzed in this study¹⁸³. Therefore, other cpn60 β genes might have substituted in folding Rubisco RbcL subunits.

Formation of the holoenzyme Rubisco complex seems furthermore dependent on specific assembly chaperones. It was reported that the product of the *rbcX* gene, present in the

intergenic space between the *rbcL* and *rbcS* genes in several cyanobacteria¹⁹⁷, enhances the production of enzymatically active Rubisco upon coexpression with *rbcL* and *rbcS* in *E.coli*¹⁹⁸⁻²⁰⁰. Partial inactivation of *rbcX* in *Synechococcus* sp. PCC7002 resulted in a substantial reduction in Rubisco solubility and activity²⁰⁰. In contrast, the *rbcX* gene was reported to be nonessential in *Synechococcus* sp. PCC7942, where the *rbcX* gene is located remote of the *rbcLS* operon¹⁹⁸.

The first crystal structure of RbcX was solved for *Synechococcus* sp. PCC7002. RbcX was shown to be dimeric of 30 kDa molecular weight and consists of four α helices per monomer that form an unusual helix bundle (Fig.17A). The core of the helical bundle is composed of conserved hydrophobic residues without authentic coiled-coil sidechain packing. The long $\alpha 4$ helices of the protomers align in an almost antiparallel fashion such that the helical bundles are located at opposite ends (Fig.17B). The $\alpha 1$ helices form additional symmetrical contacts and together with the $\alpha 2$ helices delineate a narrow diagonal groove in the arc-shaped complex. The dimer interface is predominantly uncharged and hydrophobic²⁰¹. By means of a multiple sequence alignment of cyanobacterial RbcX sequences, two highly conserved regions were identified, representing potential protein-protein interaction sites: the central groove of the dimer and a surface region around the corners of the molecule (Fig.17C). The groove in the centre of the molecule is lined by a hydrophobic area comprising the conserved residues Y17, Y20 and I50 of each monomer. Central access into the crevice is constricted by the conserved Q51 residue to a 5.4 Å wide opening that is just large enough to accommodate a polypeptide chain in an extended conformation. The groove binds to the C-terminus of RbcL that contains the recognition motif EIKFEFD and is located on the surface of RbcL subunits²⁰¹. The peptide is bound in an extended conformation with the side chains F462 and F464 occupying hydrophobic pockets of the RbcX cleft. The C-terminal recognition motif is highly conserved among all form I Rubisco homologs. Archaeal and form II Rubisco lack the consensus sequence as they also do not encode for RbcX in their genome. The other conserved region has predominantly polar character and is located at the corners of the RbcX dimer, comprising residues Q29 and R70. Because of the 2-fold symmetry, this region occurs twice at opposing edges of the dimer. As mutation of these polar interaction areas results in formation of soluble but not properly assembled RbcL in *E.coli*, it was suggested that the polar surface regions are responsible for proper subunit assembly, whereas the central groove is essential for the initial recognition and binding for the production of soluble RbcL²⁰¹. Only the combination of both interactions results in the formation of RbcL₈ core complexes. It was suggested that the initial binding to the RbcL C-terminus helps to maintain RbcL subunits in a soluble state and keeps RbcL in an orientation relative to RbcX that favors interaction with the peripheral binding surfaces²⁰¹.

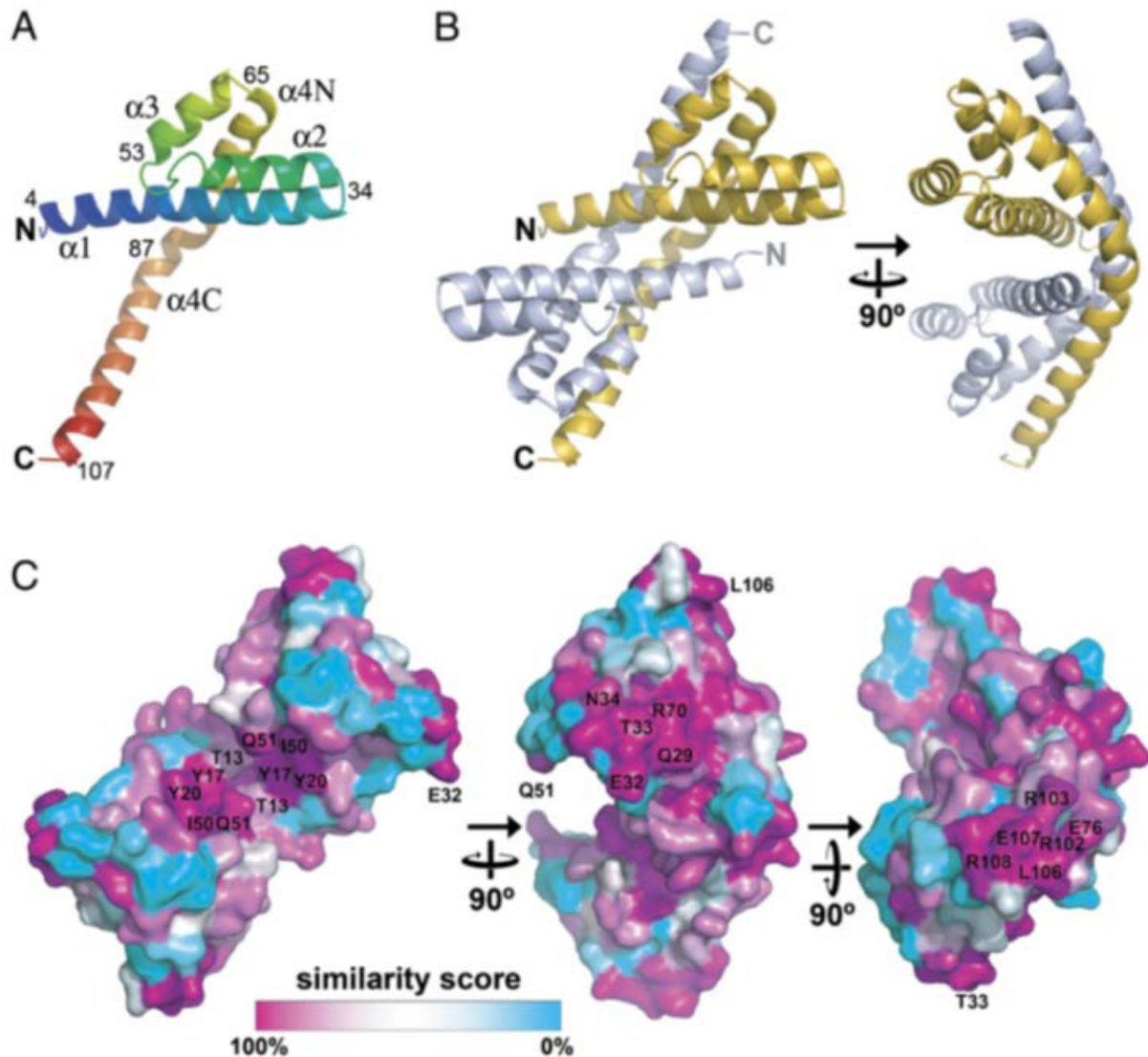


Fig.17: Crystal structure of *Synechococcus* sp. PCC 7002-RbcX. A: Ribbon representation of the RbcX monomer. The peptide backbone is depicted from N to C terminus using a color gradient from blue to red. Secondary structure elements, selected residue numbers, and chain termini are indicated. B: Structure of the RbcX dimer. Protomers are shown in yellow and blue. C: Surface conservation in RbcX. The similarity score from an alignment of 151 sequences of cyanobacterial RbcX in the PFAM database was plotted onto the accessible surface of the RbcX dimer. Sequence conservation is indicated by a color gradient, indicating highly conserved residues in magenta and variable regions in cyan. The positions of conserved surface residues are indicated. (reproduced from Saschenbrecker *et al.*²⁰¹).

It was shown, that disruption of the RbcL C-terminus abolished the formation of soluble, assembly competent RbcL protein. Thus, RbcX appears to protect this sequence from undergoing aberrant interactions. Interestingly, C-terminal truncation or mutation of RbcL also disrupted the RbcX-independent assembly of *Synechococcus* sp. PCC6301-RbcL, indicating a direct role of the C-terminus in forming or stabilizing the RbcL₈ core complex²⁰¹. The interaction between cognate RbcX and RbcL₈ has a sufficiently fast off rate to allow for efficient displacement of RbcX by RbcS, which binds stably to RbcL₈^{201,202}. The binding constant of

RbcX for the C-terminal RbcL peptide alone are in the 100 μM range, but the overall affinity of RbcX for the RbcL₈ core complex appears to be close to 1 μM . This could be attributed to the polar corner surface of RbcX, which apparently contributes to the overall binding¹⁸¹. Recently, cyanobacterial Rubisco was reconstituted *in vitro* with the assistance of a heterologous high-affinity RbcX from *Anabaena* sp. CA²⁰³. Furthermore it could be shown, that binding of RbcX to RbcL₈ is inhibiting Rubisco activity by disordering an N-terminal loop of the RbcL subunit. RbcS binding displaces RbcX by ordering this loop region rendering Rubisco catalytically competent²⁰⁴. The C-terminal RbcX recognition sequence is conserved in higher plant RbcL, and two homologs to RbcX genes are present in plant genomes exhibiting ~20-30% sequence homology to the cyanobacterial RbcX sequences. Thus, it could be possible that RbcX has an important role in the assembly process of plant Rubisco, too. Taken together, structural and functional analysis of cyanobacterial RbcX revealed that the protein acts as a homodimer by binding and stabilizing RbcL subunits subsequent to their interaction with chaperonin and assist in the efficient formation of RbcL₈ core complexes. The RbcL-RbcX interaction is dynamic, facilitating displacement of RbcX from RbcL₈ by RbcS subunits to produce the active holoenzyme (Fig.18)^{201,204}.

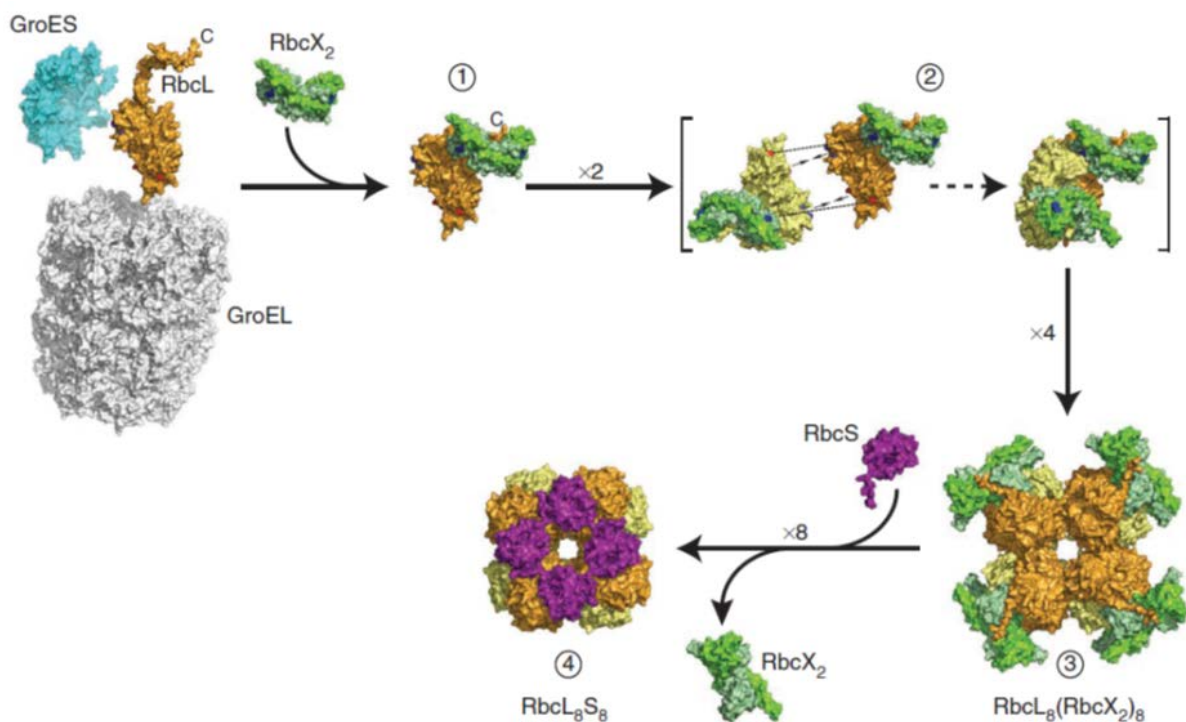


Fig.18: Model of RbcX function in cyanobacterial Rubisco assembly. 1: RbcX functions to increase the efficiency of Rubisco assembly by acting on folded RbcL subunits subsequent to their GroEL/GroES mediated folding. 2: and 3: Folded RbcL monomers may spontaneously form dimers or interact with RbcX immediately upon release from GroEL, resulting in stabilization of assembly intermediates competent for efficient progression to RbcL₈ core particles. 4: The complex between RbcL and RbcX is dynamic, facilitating the eventual displacement of RbcX by RbcS subunits to form the functional holoenzyme (reproduced from Bracher *et al.*²⁰⁴).

1.4.4 Role of auxiliary proteins in Rubisco biogenesis and function

In the following review by Hauser *et al.*²¹², most recent findings about factors involved in Rubisco folding/assembly and maintenance are described.

Role of auxiliary proteins in Rubisco biogenesis and function

Thomas Hauser[†], Leonhard Popilka[†], F. Ulrich Hartl and Manajit Hayer-Hartl^{*}

Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) catalyses the conversion of atmospheric CO₂ into organic compounds during photosynthesis. Despite its pivotal role in plant metabolism, Rubisco is an inefficient enzyme and has therefore been a key target in bioengineering efforts to improve crop yields. Much has been learnt about the complex cellular machinery involved in Rubisco assembly and metabolic repair over recent years. The simple form of Rubisco found in certain bacteria and dinoflagellates comprises two large subunits, and generally requires the chaperonin system for folding. However, the evolution of hexadecameric Rubisco, which comprises eight large and eight small subunits, from its dimeric precursor has rendered Rubisco in most plants, algae, cyanobacteria and proteobacteria dependent on an array of additional factors. These auxiliary factors include several chaperones for assembly as well as ATPases of the AAA+ family for functional maintenance. An integrated view of the pathways underlying Rubisco biogenesis and repair will pave the way for efforts to improve the enzyme with the goal of increasing crop yields.

Photosynthesis is a fundamental process in biology as it converts solar energy into chemical energy and thus, directly or indirectly, fuels nearly all life on earth. The chemical energy generated during the light reaction of photosynthesis is used to fix atmospheric CO₂ and produce reduced carbon compounds in the Calvin–Benson–Bassham cycle, the second step of the process (Fig. 1). The key enzyme responsible for carbon fixation in all photosynthetic organisms is Rubisco¹, which catalyses the conversion of an estimated ~10¹¹ tons of CO₂ per annum into organic material².

Rubisco catalyses the often rate-limiting step of photosynthetic carbon fixation in the Calvin–Benson–Bassham cycle: the carboxylation of the 5-carbon sugar substrate ribulose-1,5-bisphosphate (RuBP; Fig. 1). The resulting 6-carbon intermediate is unstable and is converted into two molecules of 3-phosphoglycerate (3PG),

which then produces glyceraldehyde-3-phosphate (G3P) via a series of steps requiring ATP and NADPH from the light reaction of photosynthesis. G3P gives rise to the synthesis of sugar molecules, fatty acids and amino acids, and is also used to regenerate RuBP (ref. 3).

Rubisco, however, is an inefficient enzyme: its catalytic rate is only ~3–10 CO₂ molecules per second (refs 4,5). Moreover, oxygen can compete with CO₂ during catalysis⁶, resulting in the production of only one molecule of 3PG and one molecule of 2-phosphoglycolate^{7,8} (Fig. 1). In plants, 2-phosphoglycolate is recycled back to 3PG via photorespiration, an energy-consuming pathway that liberates previously fixed carbon as CO₂ (ref. 9). These shortcomings may be due to Rubisco being an ancient enzyme that evolved more than 3.5 billion years ago, when the atmosphere was high in CO₂ and free of oxygen¹⁰. Additional complexity comes from the fact that the

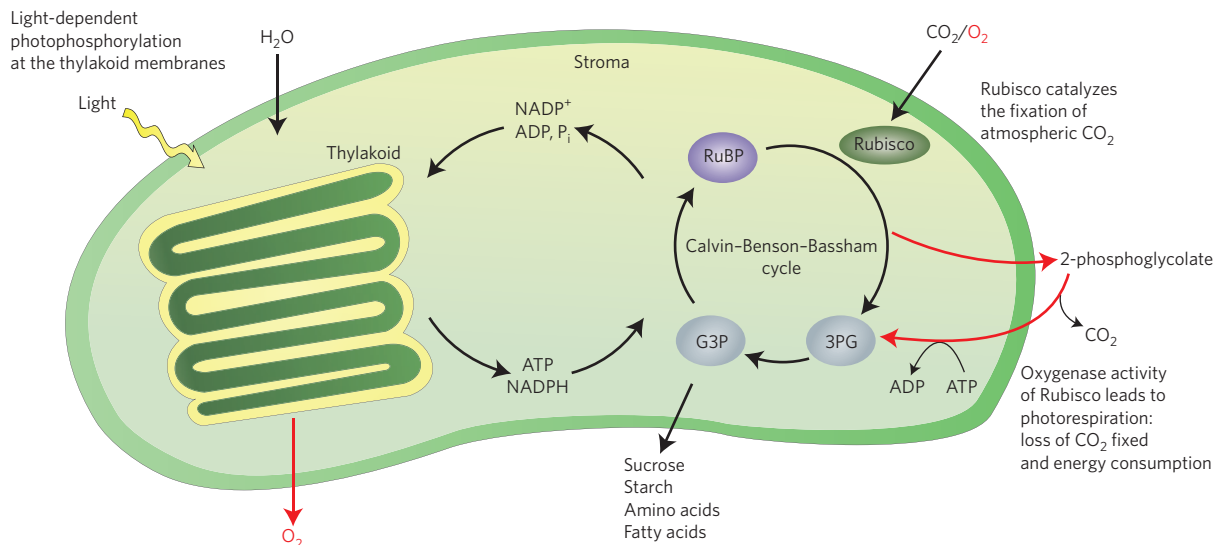


Figure 1 | Central role of Rubisco in photosynthesis. Schematic representation of photosynthesis in chloroplasts. The light reaction and the Calvin–Benson–Bassham cycle of CO₂ fixation, as well as the side-reaction of photorespiration are shown.

Department of Cellular Biochemistry, Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany.

[†]These authors contributed equally to this work. *e-mail: mhartl@biochem.mpg.de

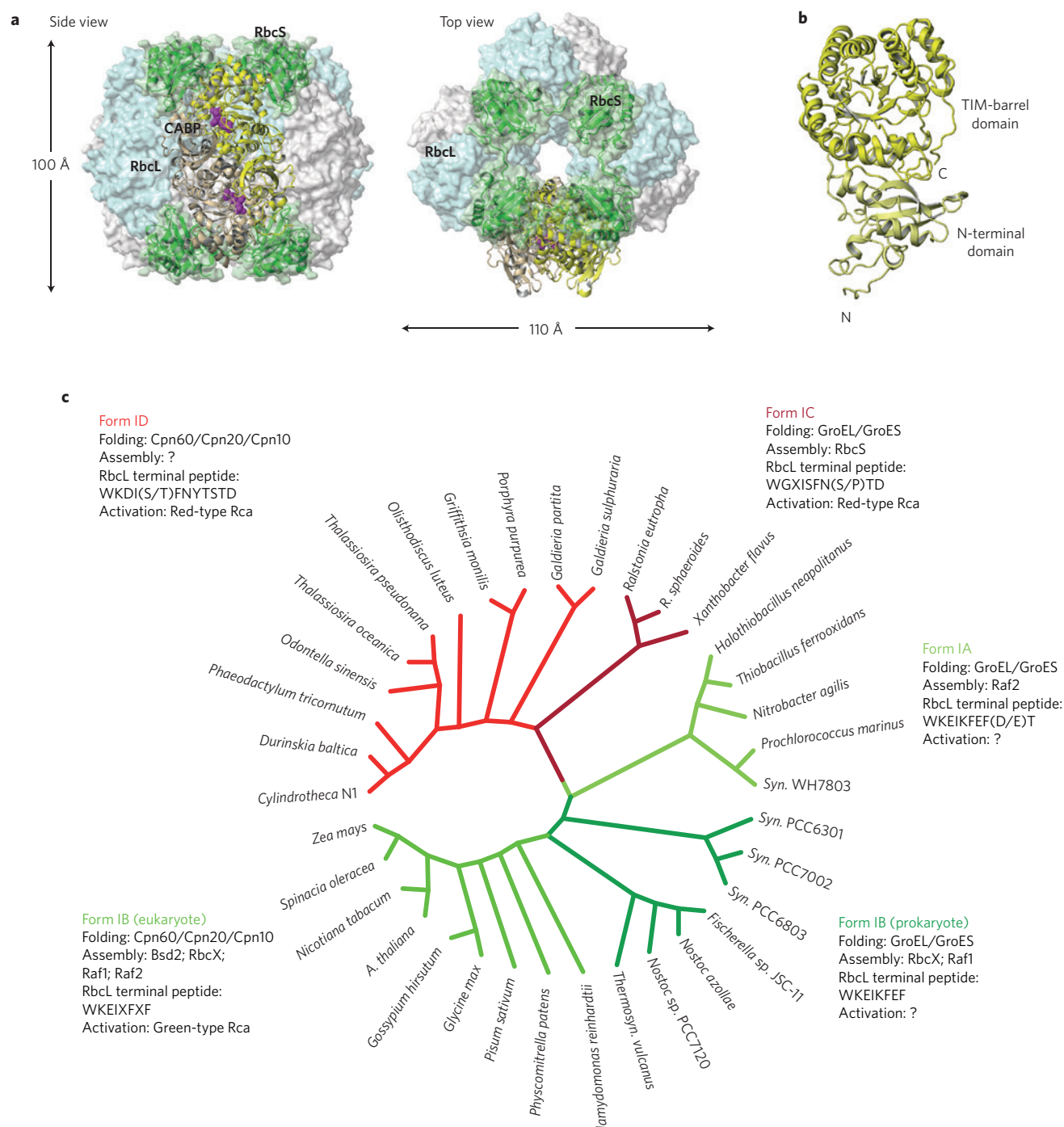


Figure 2 | Comparison of green-type and red-type Rubisco. a, Structure of hexadecameric Rubisco from the cyanobacterial species *S. elongatus* PCC6301 (Protein data bank (PDB) accession: 1RBL), shown in surface representation. One antiparallel dimer of RbcL and the adjacent RbcS are highlighted in ribbon representation. The transition state analogue carboxyarabinitol-1,5-bisphosphate (CABP) is bound in the active sites. **b**, Structure of a single RbcL subunit in ribbon representation, showing the N-terminal $\alpha\beta$ domain and the C-terminal TIM-barrel domain. **c**, Molecular phylogenetic tree of selected Rubisco RbcL sequences (modified from ref. 16, Oxford Journals). The green-type enzymes encompass form IA and IB and the red-type form IC and ID. Components associated in Rubisco folding, assembly and activation identified in the different clades are indicated. The phylogenetic tree was calculated with T-coffee (ref. 99) by multi sequence alignment and the diagram created using the software Dendroscope (ref. 100).

multistep catalytic reaction is prone to occasional processing errors that produce ‘misfire’ products, some of which can remain bound to Rubisco, inhibiting catalysis¹¹.

Recent forecasts suggest that global food production will need to rise by more than 25% by 2050 to meet the ever increasing demand¹². Engineering a catalytically more-efficient Rubisco enzyme could

contribute to reaching that goal, but the complex nature of Rubisco’s folding and assembly has made these efforts exceedingly challenging^{10,13–15}. Here we review recent progress in understanding the complex cellular chaperone machineries that ensure the efficient biogenesis of this most abundant enzyme and maintain it in a functionally active state.

Rubisco structure and catalytic function

Several structural forms of Rubisco are found in nature¹⁶. The most common — form I — in plants, algae, cyanobacteria and proteobacteria has a hexadecameric structure consisting of eight large (RbcL, ~50 kDa) and eight small (RbcS, ~15 kDa) subunits¹. The RbcL subunits arrange as a tetramer of antiparallel dimers (the RbcL₈ core complex), capped by four RbcS subunits at the top and four at the bottom. The ~520 kDa holoenzyme has a cylindrical shape with a diameter of ~110 Å and a height of ~100 Å (Fig. 2a). The simpler form II Rubisco found in certain bacteria and dinoflagellates consists only of a dimer of RbcL subunits with no RbcS.

The overall structure of the RbcL subunits is similar in all forms of Rubisco, and consists of a N-terminal $\alpha\beta$ domain of ~150 amino acids and a C-terminal $\beta_8\alpha_8$ triosephosphate isomerase (TIM)-barrel domain of ~320 amino acids¹ (Fig. 2b). Each antiparallel RbcL dimer has two active sites formed by residues in the N-domain of one subunit and the C-domain of the adjacent subunit (Fig. 2a). Although their RbcL sequences are highly conserved (~80% amino acid identity¹⁷), form I Rubiscos are phylogenetically classified into green-type enzymes (forms IA and IB of cyanobacteria, green algae and plants) and red-type enzymes (forms IC and ID of non-green algae and phototrophic bacteria)¹⁶ (Fig. 2c). The RbcS subunits are more diverse in sequence, generally with only 30–40% homology between different species. Their common core structure consists of a four-stranded antiparallel β -sheet covered on one side by two helices¹⁸ (Fig. 2a). The most significant variation occurs in two distinct locations, the loop between β -strands A and B, and the C-terminal region which is elongated in the enzymes of red-type organisms (Supplementary Fig. 1). The diversity in small subunit structure appears to significantly contribute to functional differences between Rubisco enzymes^{18,19}. In eukaryotes the RbcL subunits are always plastid encoded, whereas the RbcS subunits are plastid encoded only in non-green algae and nuclear encoded in plants and green algae²⁰. The nuclear-encoded RbcS subunits carry cleavable N-terminal targeting sequences for import into chloroplasts²¹. Notably, some Rubiscos of red algae and diatoms are catalytically more efficient than the green-type enzymes²², making them attractive targets for plant biotechnology^{10,15}.

To acquire catalytic competence, Rubisco must first undergo an activation process. In this reaction, the ϵ -amino group of a highly conserved lysine (K201 in spinach RbcL) in the active site is carbamylated by a non-substrate CO₂ molecule, followed by the binding of Mg²⁺ to the carbamate moiety⁶. This initial carbamylation of the active site pocket is essential to position the RuBP for efficient electrophilic attack by the substrate CO₂ molecule. Upon binding of RuBP, the so-called loop 6 of the TIM-barrel domain covers the active site and finally the flexible C-terminal strand of the RbcL subunit fixes loop 6 in position via multiple polar interactions²³. As a result, there is no solvent access to the active centre and catalysis can proceed. As will be discussed below, the activation and catalytic reaction are error prone, generating inactive enzyme complexes that must be repaired by the chaperone Rubisco activase.

Folding of the Rubisco large subunit by chaperonin

More than 30 years ago, experiments in chloroplasts showed that newly synthesized RbcL subunits interact with a large protein complex prior to forming the holoenzyme, suggesting that this 'Rubisco large subunit binding protein' was involved in the assembly process²⁴. The Rubisco subunit binding protein was later identified to be the chloroplast homologue of bacterial GroEL and mitochondrial Hsp60, a class of molecular chaperones referred to as chaperonins^{25,26}. These ATP-regulated macromolecular machines function as nano-compartments for single protein molecules to fold in isolation, unimpacted by aggregation. Indeed, it was shown more recently that the chaperonin system mediates the folding of the RbcL subunits, not their assembly^{27,28}.

The chaperonins are large cylindrical complexes consisting of two heptameric rings of ~60 kDa subunits stacked back to back²⁶. Whereas bacterial GroEL and mitochondrial Hsp60 are homooligomeric, the chloroplast chaperonin Cpn60 is composed of homologous α and β subunits²⁹. These chaperonins cooperate functionally with single ring cofactors that bind transiently to the ends of the cylinder: GroES in bacteria; Hsp10 in mitochondria; Cpn10 and Cpn20 in chloroplasts³⁰. While GroES, Hsp10 and Cpn10 are heptamers of ~10 kDa subunits, Cpn20 is a tandem repeat of Cpn10 units that may function alone or as a hetero-oligomer with Cpn10 (refs 31,32). The special features of the chloroplast chaperonin system may represent an adaptation to chloroplast specific substrate proteins.

The best studied chaperonin system is the bacterial GroEL and its cofactor GroES (Fig. 3a). Each subunit of GroEL is composed of an apical domain, an intermediate hinge domain and an equatorial ATPase domain (Fig. 3b). The apical domains form the flexible ring opening and expose hydrophobic amino acids for the binding of non-native substrate protein and GroES. The ATP-regulated binding of GroES results in the displacement of the substrate protein into an enclosed cage, large enough to accommodate proteins up to ~60 kDa (refs 26,33; Fig. 3c). Concomitant with GroES binding to the GroEL ring that holds the substrate, the GroEL subunits undergo conformational changes that render the lining of the central cavity hydrophilic and thus permissive for substrate protein folding. The enclosed protein is free to fold for ~2–7 seconds (dependent on temperature), the time needed for ATP hydrolysis. Subsequent binding of ATP to the opposite GroEL ring causes GroES dissociation and substrate release. Proteins that are not yet stably folded are rapidly recaptured for another folding attempt (Fig. 3c). The GroEL–GroES system of bacteria and the chloroplast chaperonin appear to function by essentially the same mechanism^{31,32,34}.

Many of the obligate substrates of *Escherichia coli* GroEL–GroES have structural similarity to RbcL in that they also contain the TIM-barrel domain³⁵. These proteins are generally highly aggregation prone due to their tendency to populate kinetically trapped folding intermediates with exposed hydrophobic amino acids. They must fold inside the GroEL–GroES cage to avoid aggregation and reach the native state efficiently. The RbcL subunits of bacterial form II Rubisco are also highly aggregation prone but it was found that active enzyme could be obtained in the presence of GroEL, GroES and ATP *in vitro*³⁶. It was later shown that the RbcL subunit requires encapsulation in the GroEL–GroES cavity for folding and then assembles spontaneously to the active dimer²⁷. Renaturation experiments under conditions avoiding aggregation (low RbcL concentration and temperature) revealed that the spontaneous folding of RbcL is slow but is accelerated ~3-fold by the chaperonin²⁷. Similar observations were made with other TIM-barrel substrate proteins, where the rate of folding in the presence of GroEL–GroES is accelerated between 30- and 100-fold³⁷. Recent efforts to reconstitute form I Rubisco *in vitro* showed that the form I RbcL also requires the chaperonin for folding, but fails to assemble and remains chaperonin associated²⁸. The failure of the folded RbcL subunit to assemble under *in vitro* conditions suggests that it retains substantial structural flexibility upon transient release from the chaperonin²⁸.

RbcX, an assembly chaperone for some green-type Rubiscos

Assembly of the RbcL₈S₈ holoenzyme is generally thought to involve the formation of the RbcL₈ core complex followed by the docking of unassembled RbcS subunits, which can fold spontaneously *in vitro*. Various form I cyanobacterial Rubiscos have been expressed in *E. coli*, generally with low yields of 0.1–10%³⁸, although coexpression of GroEL–GroES improves the yield in the case of *Synechococcus elongatus* PCC6301 (Syn6301)^{36,39}. Notably, the Rubiscos from Syn6301 and *Thermosynechococcus elongatus* BP-1 are the only cyanobacterial proteins so far for which a stable

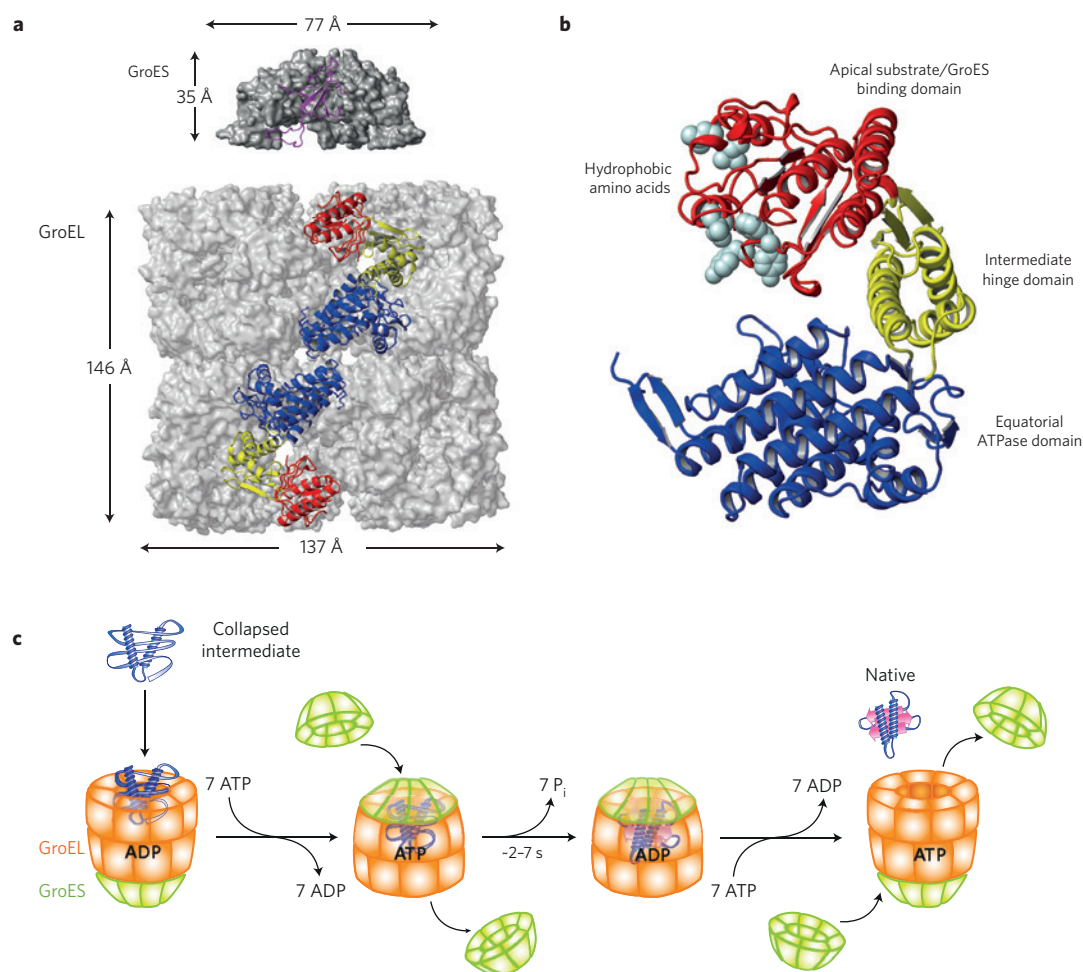


Figure 3 | The bacterial chaperonin system. **a**, Structure of the GroEL tetradecamer complex (PDB: 1GRL) and the GroES heptamer (PDB: 1AON) in surface representation. Two subunits of GroEL in opposite rings and one subunit in GroES are highlighted in ribbon representation. **b**, Single subunit of GroEL in ribbon representation, indicating the apical, intermediate and equatorial domains. The grey spheres represent hydrophobic residues that point towards the central cavity of the GroEL cylinder and mediate binding of non-native substrate protein. **c**, The generic GroEL-GroES reaction cycle. Substrate protein as a collapsed folding intermediate is bound by the open GroEL ring of the asymmetrical GroEL-GroES complex. Binding of ATP to the substrate-bound ring causes a conformational change in the apical domains which results in the exposure of the GroES binding residues. Binding of GroES causes substrate displacement into an enclosed folding cage. ADP and GroES dissociate from the opposite ring together with previously-bound substrate (not shown). The newly-encapsulated substrate is free to fold in the GroEL cavity during the time needed to hydrolyse the bound ATP molecules (~2-7 s dependent on temperature). ATP binding followed by GroES binding to the opposite ring triggers GroES-dissociation, releasing the substrate protein. Figure reproduced from ref. 26, Annual Reviews.

RbcL₈ complex could be produced in the absence of RbcS (refs 40–42). These findings suggested that additional factors, besides the chaperonin, are necessary to ensure the efficient formation of the structurally more complex form I Rubisco. One possible candidate was the protein encoded by the *rbcX* gene, which is located in the intergenic space between the *rbcL* and *rbcS* genes of several cyanobacterial species³⁸ and is conserved in all organisms with form IB Rubisco, including plants⁴³ (Fig. 2c). Coexpression of RbcX with RbcL and RbcS in *E. coli* enhanced the production of enzymatically active Rubisco^{39,41,44,45}. Partial inactivation of the *rbcX* gene in *Synechococcus* sp. PCC7002 (Syn7002) resulted in a substantial reduction in Rubisco solubility and activity⁴⁵, while in *S. elongatus* PCC7942 (Syn7942) the *rbcX* gene was shown to be nonessential³⁹. Although the RbcX from *Arabidopsis thaliana* is functional when substituted into *E. coli* expressing cyanobacterial Rubisco⁴⁶, a requirement of RbcX in Rubisco biogenesis in plants remains to be established.

RbcX functions as a homodimer of ~15 kDa subunits^{41,47–49}. Recent biochemical and structural analyses demonstrated its role as a specific assembly chaperone. Coexpression of RbcX in *E. coli* was essential for the soluble expression of Syn7002-RbcL and complexes between RbcL and RbcX assembly intermediates were detected by co-immunoprecipitation⁴¹. A direct interaction between RbcL and RbcX was also detected in chloroplasts^{46,49}. The crystal structure of Syn7002-RbcX revealed that each subunit consists of four α -helices that form a helical bundle at one end. The two long α 4 helices of each protomer align in an almost antiparallel fashion, with a ~60° kink midway, resulting in the helix bundles being located at opposite ends of the boomerang-shaped dimer (Fig. 4a). A highly conserved diagonal groove in the centre of the dimer is large enough to accommodate a peptide in an extended conformation (Fig. 4b). Similar structures were shown for the RbcX proteins of the thermophilic cyanobacterium *T. elongatus* BP-1 and for both RbcX homologues of *A. thaliana*^{48,49}. The central groove is hydrophobic

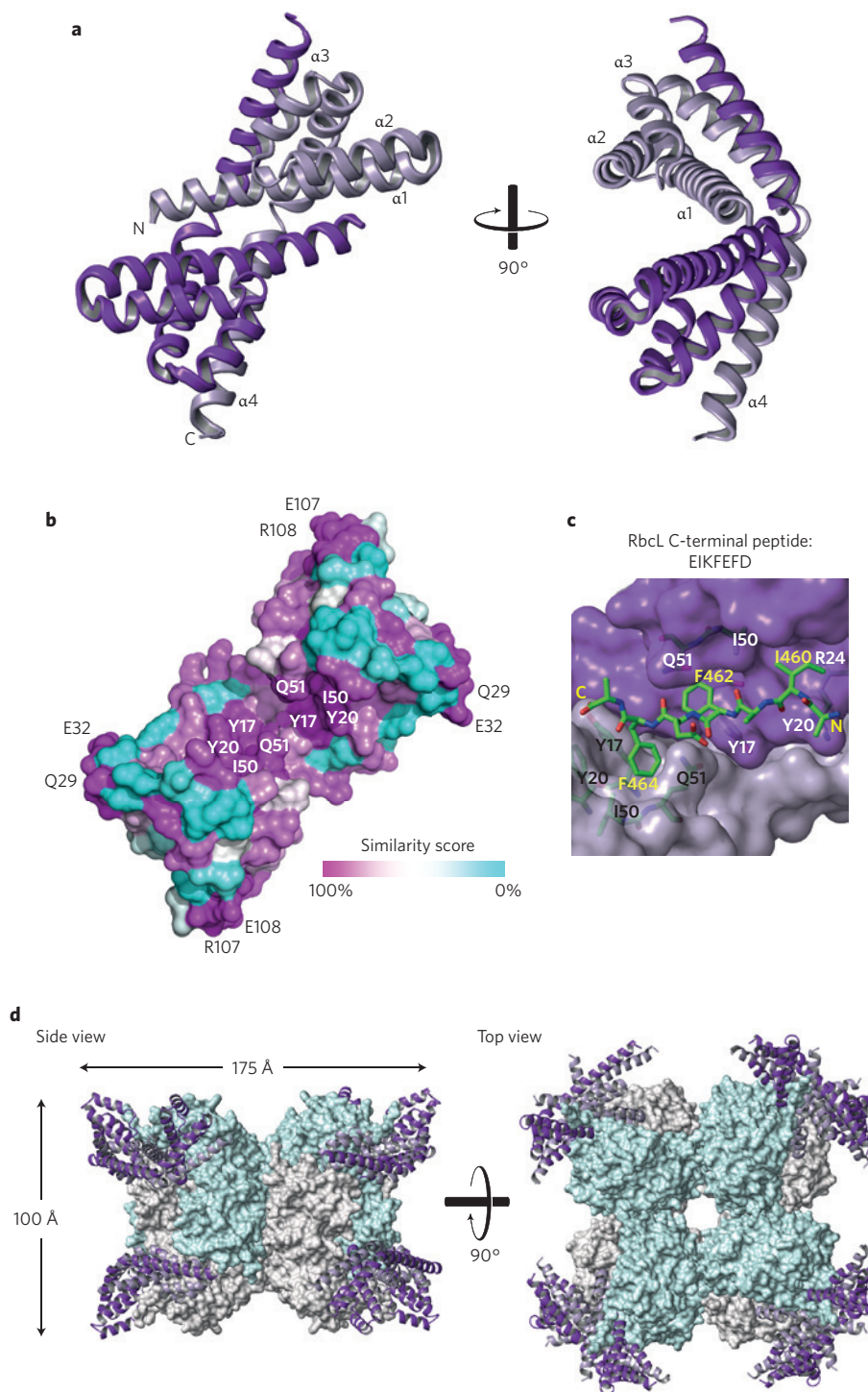


Figure 4 | The RbcX assembly chaperone. **a**, Structure of the RbcX dimer from the cyanobacterial species Syn7002 (PDB: 2PEI). Protomers are shown in ribbon representation. **b**, Surface conservation of RbcX. The similarity score from an alignment of 151 sequences of cyanobacterial RbcX in the PFAM database was plotted onto the accessible surface of the RbcX dimer. Sequence conservation is displayed as a color gradient, indicating highly conserved residues in magenta and variable regions in cyan. The positions of conserved surface residues are indicated. **c**, Peptide binding cleft of RbcX. The C-terminal peptide EIKFEFD of RbcL, shown in ball-and-stick representation, is bound in the central cleft of the RbcX, shown in surface representation (PDB: 2PEM). The positions of residues in RbcX critical for peptide binding as well as the critical residues of the peptide are indicated. N- and C-termini of the peptide are also indicated. **d**, Overall architecture and dimensions of the RbcL₈RbcX₈ assembly intermediate. The RbcL₈ core structure of Syn6301 is shown in surface representation and the bound RbcX in ribbon representation. The C-terminal tails of RbcL subunits are bound within the central cleft of RbcX and the conserved corner residues of RbcX are in contact with the N-domain of the adjacent RbcL subunit in the antiparallel dimer. **a–c**, modified from ref. 41, Cell Press. **d**, modified from ref. 50, NPG.

and binds the sequence motif EIKFEFD present at the C-terminus of RbcL subunits of all form IB Rubiscos (Fig. 4c). Polar residues at the four corners of the RbcX dimer are also highly conserved (Fig. 4b), and mutations in these regions as well as in the peptide binding cleft abolished function⁴¹. While the interaction between RbcL and RbcX is normally dynamic, the use of a heterologous system (Syn6301-RbcL and RbcX from *Anabaena* sp. CA) allowed the formation of a stable complex consisting of RbcL₈ and eight RbcX dimers⁵⁰ (Fig. 4d). In the crystal structure of this ~670-kDa complex, each antiparallel RbcL dimer has two RbcX dimers bound, one on top and one at the bottom. The C-terminal tail of the RbcL chain is bound within the central cleft of RbcX, in an extended conformation, while conserved polar residues from the corner regions of RbcX form critical hydrogen bonds and salt bridges with residues in the N-terminal domain of the adjacent RbcL chain⁵⁰ (Fig. 4d). Thus, the two RbcX dimers act as 'molecular clamps' that stabilize the RbcL dimer. Interestingly, the N-terminal 18-amino-acid residues and residues 64–70 of the so-called 60s-loop of RbcL are disordered in the RbcL₈RbcX₈ complex. In contrast, these residues are structured in the holoenzyme by interactions within the RbcL dimer and with RbcS. However, the position of the 60s-loop in the holoenzyme would be sterically incompatible with RbcX binding, and structure formation induced by RbcS would cause the displacement of RbcX (ref. 50). This conformational change in RbcL upon RbcS binding provides an explanation for why RbcS is required for catalytic activity^{50,51}.

The cyanobacterial form I Rubisco could be reconstituted *in vitro* using GroEL–GroES and RbcX (ref. 28). These experiments revealed the requirement of RbcX for stabilization of RbcL in a form that no longer rebinds to GroEL, consistent with findings that RbcX can mediate RbcL₂ and RbcL₈ formation^{41,52}. Together, these data suggest the following model for RbcX function (Fig. 5a): after folding and release by GroEL/ES, the flexible C-terminus of folded RbcL binds within the central cleft of RbcX; charge pairs in the RbcL dimer interface and between RbcL and RbcX then mediate the proper antiparallel alignment of the RbcL subunits with RbcX functioning as a molecular clamp; stabilization of the RbcL dimer then shifts the equilibrium towards the formation of the RbcL₈RbcX₈ assembly intermediate; finally, RbcS binding between the RbcL dimers triggers conformational changes in the RbcL subunits, resulting in the displacement of RbcX and formation of the hexadecameric Rubisco.

Other factors involved in green-type Rubisco biogenesis

Newly-synthesized polypeptides generally interact first with chaperones that bind to the nascent chain on the ribosome, preventing misfolding and aggregation⁵³. These chaperones include the Hsp70 system, which assists in the folding of a subset of proteins through cycles of binding and release. Proteins that are unable to reach their native state by this mechanism may be transferred to chaperonin for final folding²⁶. The bacterial cytosol and the chloroplast stroma contain homologous Hsp70 machineries that are likely to act on RbcL upstream of the chaperonin⁵⁴. Likewise, the RbcS subunit utilizes these chaperones for folding after import into the chloroplast⁵⁵. Another component with a putative role in the initial stages of Rubisco biogenesis is the chloroplast-specific bundle sheath defective2 protein (Bsd2), present in both bundle sheath and mesophyll chloroplasts⁵⁶. The *bsd2* gene is required for Rubisco accumulation in maize and Bsd2 was found to cofractionate with polyribosomes synthesizing RbcL (refs 57,58). Bsd2 contains the zinc-finger motif present in several Hsp40 proteins, a family of Hsp70 cofactors^{57,59}, suggesting a functional cooperation with Hsp70.

A recent screen of a photosynthetic mutant library in maize identified several mutants with reduced Rubisco content but near normal levels of other photosynthetic enzyme complexes⁶⁰. The genes responsible were proposed to encode so-called Rubisco

accumulation factors (Raf). Two of these, Raf1 (~50 kDa) and Raf2 (~10–18 kDa), have been biochemically and structurally studied^{42,60–63}. Raf1 appears to function as a dimer^{42,63} and is conserved in plants and cyanobacteria that express form IB Rubisco (Fig. 2c). In the absence of Raf1, the amount of Rubisco holoenzyme was found to be substantially reduced, although transcription and translation of RbcL and RbcS was normal^{60,63}. Instead, RbcL subunits were associated with the chaperonin, suggesting that Raf1 acts downstream of chaperonin-assisted RbcL folding⁶⁰. Coexpression in *E. coli* of Rubisco and Raf1 from *T. elongatus* was shown to promote Rubisco assembly, apparently involving the formation of a RbcL–Raf1 assembly intermediate⁴². Moreover, coexpression of *A. thaliana* Raf1 improved the biogenesis of a hybrid Rubisco in tobacco plants⁶³. These findings classify Raf1 as a Rubisco assembly chaperone that may either act in concert with, or possibly in place of, RbcX.

Raf2 has been identified in some organisms containing form IB Rubisco and also in organisms containing form IA Rubisco in α -carboxysomes^{61,62} (Fig. 2c). The crystal structure of Raf2 from the chemotrophic bacterium *Thiomonas intermedia* K12 showed that the protein forms a homodimer of ~10-kDa subunits with homology to pterin-4 α -carbinolamine dehydratase (PCD) enzymes⁶¹. The overall PCD fold is preserved in Raf2 but the active site cleft present in PCD enzymes is disrupted⁶¹. Loss of Raf2 function in maize resulted in a weaker phenotype than disruption of the *raf1* gene⁶². Chemical crosslinking followed by immunoprecipitation suggested that Raf2 interacts with imported RbcS subunits and to a lesser extent with RbcL in the chloroplast stroma⁶². It has been proposed that Raf1, Raf2 and Bsd2 form transient complexes with RbcS that maintain RbcS competence for assembly with RbcL⁶². It may be relevant to note that the plant Raf2 contains an additional domain, not present in cyanobacterial Raf2, which may be involved in mediating these interactions. Further biochemical analysis is needed to elucidate the exact role of Raf1, Raf2 and Bsd2 in Rubisco assembly and their possible functional relationship, if any, with RbcX and other Raf proteins yet to be analysed.

Assembly of red-type Rubiscos

The Rubiscos of red-type organisms comprise the enzymes with the highest CO₂:O₂ specificity found in nature¹⁵ and thus are of considerable biotechnological importance. Interestingly, the genomes of red-type organisms do not contain sequence homologues of the assembly factors RbcX, Bsd2, Raf1 or Raf2, pointing to an alternative mechanism of Rubisco assembly. Indeed, upon coexpression in *E. coli*, the RbcL and RbcS subunits from the red-type proteobacterium *Rhodobacter sphaeroides* assembled to the holoenzyme with very high efficiency⁶⁴. RbcL subunits, when expressed alone, failed to form RbcL₈, instead populating mostly lower oligomeric states. Addition of purified RbcS to cell lysates gave rise to holoenzyme formation, suggesting a specific role of RbcS in the assembly process. *In vitro* reconstitution experiments showed that the red-type RbcL subunits also require the GroEL–GroES chaperonin system for folding. However, unlike green-type RbcL, the *R. sphaeroides* RbcL assembled efficiently in an RbcS-mediated mechanism⁶⁴. This function of RbcS to mediate assembly critically depends on the C-terminal β -hairpin sequence found exclusively in red-type RbcS proteins (Supplementary Fig. 1). The β -hairpins form β -barrel structures at the top and bottom of the holoenzyme solvent channel, serving as an assembly platform for RbcL subunits (Fig. 5b). Moreover, the red-type RbcS was able to assemble with green-type Syn6301-RbcL *in vitro*, replacing the requirement for RbcX, although the resulting heterologous complex was essentially inactive⁶⁴. Why the green-type enzymes have evolved a dependence on extrinsic assembly factors rather than using this simpler RbcS-mediated mechanism of assembly is unclear. The fact that in red-type organisms both the RbcL and RbcS subunits are expressed from the same operon (either in the bacterial cytosol or the chloroplast

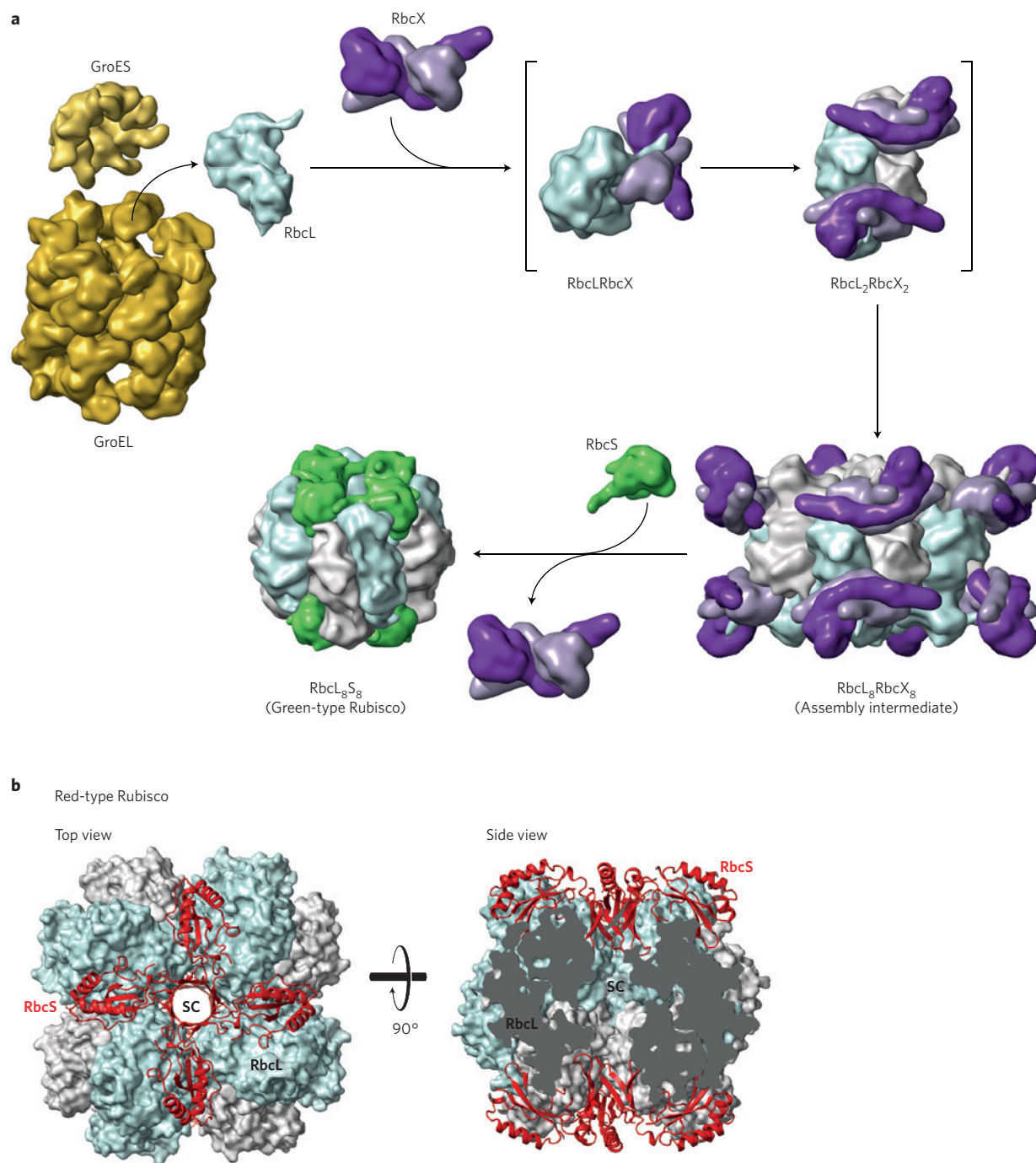


Figure 5 | Models of Rubisco assembly. a, Role of GroEL–GroES chaperonin and RbcX in green-type Rubisco assembly based on *in vitro* reconstitution and available structural information. Upon folding and release from the chaperonin complex, the cyanobacterial RbcL subunit is recognized by RbcX, which binds the flexible C-terminal RbcL peptide. Formation of the antiparallel RbcL dimer occurs mediated by two RbcX acting as molecular staples. The stable RbcL₂RbcX₂ units subsequently assemble to the RbcL₈RbcX₈ complex, in which a large portion of the RbcS binding interface is pre-formed. RbcS binding structures the RbcL N-terminus and the 60s loop, causing displacement of RbcX and formation of the functional Rubisco holoenzyme. Note that additional assembly factors, such as Raf1, are also involved in assembly but their mechanism of action remains to be determined. Figure modified from ref. 50, NPG. **b**, Role of RbcS in red-type Rubisco assembly based on mutational analyses and *in vitro* reconstitution⁶⁴. The RbcS subunit arrangement in the crystal structure of the red-type Rubisco from *Alcaligenes eutrophus* (PDB: 1BXN) is shown with the RbcS subunits in ribbon representation (red) and the RbcL octameric core in surface representation. The side view is a cross-section along the four-fold axis through the complex showing the central barrel formed by the β -hairpin extensions (see Supplementary Fig. 1) of the RbcS subunits which mediate assembly. SC, solvent channel. Figure modified from ref. 64, ASBMB.

stroma), and are thus always present simultaneously, may be important in this respect²⁰.

Maintenance of enzyme function by Rubisco activase

Once assembled, the Rubisco holoenzyme in leaves can be stable for weeks to months⁶⁵, but its activity may become compromised, necessitating Rubisco activase (Rca) as an additional chaperone for functional maintenance^{66,67}. During Rubisco's multistep catalytic reaction, protonation and oxygenation of the RuBP enediolate intermediate can result in the formation of isomeric pentulose bisphosphates, so-called misfire by-products⁶⁸. These include xylulose-1,5-bisphosphate (XuBP), 2,3-pentodiulose-1,5-bisphosphate (PDBP) and 3-ketoarabinitol-1,5-bisphosphate (KABP)^{11,68}, whose formation results in an inactive, 'closed' enzyme that reactivates only slowly, limited by the spontaneous opening of the active site (Fig. 6a). Some plants also produce a night-time inhibitor of Rubisco, 2-carboxy-D-arabinitol 1-phosphate (CA1P). Moreover, inactive Rubisco can also be generated during the error-prone activation of the catalytic centre when the substrate RuBP binds before carbamylation of lysine 201 has occurred⁶ (Fig. 6a). In all these cases, Rubisco must be conformationally remodelled by Rca in order to release inhibitory sugar molecules (Fig. 6a).

Rca belongs to a subgroup of ATPases, AAA, associated with various cellular activities, referred to as AAA+. Many of these proteins form hexameric complexes with a central pore^{69–71}. They can act as 'unfolds' by using the energy of ATP to exert mechanical force on their target proteins, typically by pulling extended sequences or loop segments into the central pore. Recent crystal structures showed that Rca from green- and red-type organisms share the AAA+ domain architecture consisting of an N-terminal α/β nucleotide binding subdomain and a C-terminal α -helical subdomain^{72–74} (Fig. 6b,d). In the case of red-type Rca from *R. sphaeroides* (also known as CbbX), formation of the active hexamer requires ATP and allosteric regulation by the binding of RuBP to a conserved pocket in the C-terminal α -helical subdomain⁷² (Fig. 6b). Regulation by RuBP, the substrate of Rubisco, ensures that Rca forms the hexamer complex when photosynthesis is active. The disc-shaped hexamer docks onto Rubisco with its highly-conserved top surface. This positions Rca to engage the extended, flexible C-terminus of an RbcL subunit with its narrow (~25 Å) (Fig. 6c) central pore, resulting in a ~4-fold stimulation of the Rca ATPase activity⁷². Reversible threading of the C-terminus of RbcL into the central pore is mediated by a classical pore-loop sequence (aromatic-hydrophobic-glycine) present in many AAA+ proteins⁶⁹ and facilitates the opening of the Rubisco active site, allowing release of the inhibitory sugar⁷². In the case of XuBP, a specific hydrolase, CbbY, cooperates with the action of Rca by degrading the released inhibitor, thereby preventing its rebinding⁷⁵, an interesting example of metabolite damage repair⁷⁶. The genomes of photosynthetic organisms from bacteria to plants all contain CbbY homologues⁷⁵.

The mechanism of Rca in plants and other green-type organisms is less well understood. Oligomeric complexes varying in size have been observed^{77–82}, although the hexamer is the likely functional state⁷³. The central pore of tobacco Rca is wider (~36 Å) than that of *R. sphaeroides* Rca and lacks the classical pore-loop residues^{72,73}, suggesting that it might act in remodelling Rubisco by a different mechanism. In support of this possibility, green-type RbcL lacks the extended C-terminus of the red-type. Moreover, green-type Rca is constitutively ATPase active⁶⁶ and has no binding site for RuBP. It is instead regulated by the intracellular ATP/ADP-ratio and by reversible disulphide-bond formation in the C-terminal extensions of larger isoforms found in some plants⁶⁶. Substrate recognition by green-type Rca seems to be more complex, involving an additional N-domain not present in red-type Rca as well as regions in the α -helical subdomain (helix 9). In tobacco, these regions interact with residues 89 and 94 (spinach numbering) of RbcL subunits located

in the equatorial region of RbcL_{8S} and confer the ability of Rca to discriminate between solanaceous and non-solanaceous Rubisco^{66,83} (Fig. 6d,e). Accordingly, green-type Rca only activates green-type Rubisco and red-type Rca is specific for red-type Rubisco. In fact, the low sequence similarity and the different regulation of the two groups of activase suggests that these enzymes have emerged independently by convergent evolution. With regard to plant bioengineering, overcoming the inherent thermal lability of green-type Rca is considered a possible strategy for improving crop plant viability under the higher temperatures predicted^{15,84}.

Evolutionary considerations and perspectives

Considering the dependence of photosynthetic organisms on functional Rubisco for growth and survival, and the significant energetic investment required to produce large quantities of the enzyme, it is surprising that evolution has failed to further enhance Rubisco catalytic efficiency. It has been suggested that Rubisco is residing in an evolutionary trap, having initially evolved in an oxygen-free atmosphere^{10,85,86}. A further consideration is that the extensive dependence of Rubisco on auxiliary factors for folding, assembly and maintenance may have impacted its ability to escape from its evolutionary trapped state¹⁵. Molecular chaperones that assist the folding of a range of client proteins, such as the chaperonins, are generally thought to promote the structural evolution of proteins by buffering deleterious effects of mutations on foldability and stability^{87–91}. In the bacterial system, overexpression of GroEL–GroES allowed monomeric proteins to accumulate an increasing number of destabilizing mutations^{92,93}. A similar effect of GroEL/ES, albeit less significant, was observed in recent studies of Rubisco evolvability^{94–96}. The more limited buffering effect of the chaperonin in this case was attributed to the fact that the wild-type RbcL subunit is already chaperonin-dependent⁹⁶, consistent with the notion that adaptation to obligatory chaperone dependence may slow evolvability⁹⁰. Interestingly, the specific assembly chaperone RbcX restricted the range of mutations compatible with functionality and prevented a subset of mutants from forming the holoenzyme⁹⁶. Thus, RbcX appears to reinforce the evolutionary trapped state of Rubisco, with other assembly chaperones presumably having a similar effect. Nevertheless, using the Syn6301-Rubisco, which can assemble without RbcX, it proved possible to select a mutation in RbcL with improved catalytic properties⁹⁶. However, this approach allows one to sample only a minute fraction of the available mutational space, particularly when considering that auxiliary factors may have to co-evolved in order to express functionally improved Rubisco variants. Indeed, recent evidence showed Rubisco biogenesis in leaves is dependent on the evolutionary compatibility between Rbf1 and RbcL (ref. 63).

Given these challenges, other options to improve Rubisco performance in crop plants include the engineering of cyanobacterial CO₂-concentrating mechanisms^{97,98} and the expression of a catalytically superior red-type Rubisco in chloroplasts¹⁵. The latter approach, although not successful so far²², seems promising in light of the finding that red-type Rubiscos may be less dependent on auxiliary factors for assembly. However, in this case, the introduction of a compatible Rca with the red-type Rubisco may be necessary for optimal activity.

A better mechanistic understanding of Rubisco biogenesis, activity regulation and the chaperone machineries involved may hold promise for new bioengineering strategies based on a rational design. Importantly, despite considerable efforts, it has not yet been possible to recombinantly produce plant Rubisco in *E. coli* or to reconstitute the holoenzyme *in vitro*. The re-invigorated interest in Rubisco has discovered, and will continue to unearth, exciting new features of the enzyme that will aid our capacity to modify its biogenesis and activity in a range of photosynthetic organisms — especially those tailored for addressing the challenges of improving food and fuel production.

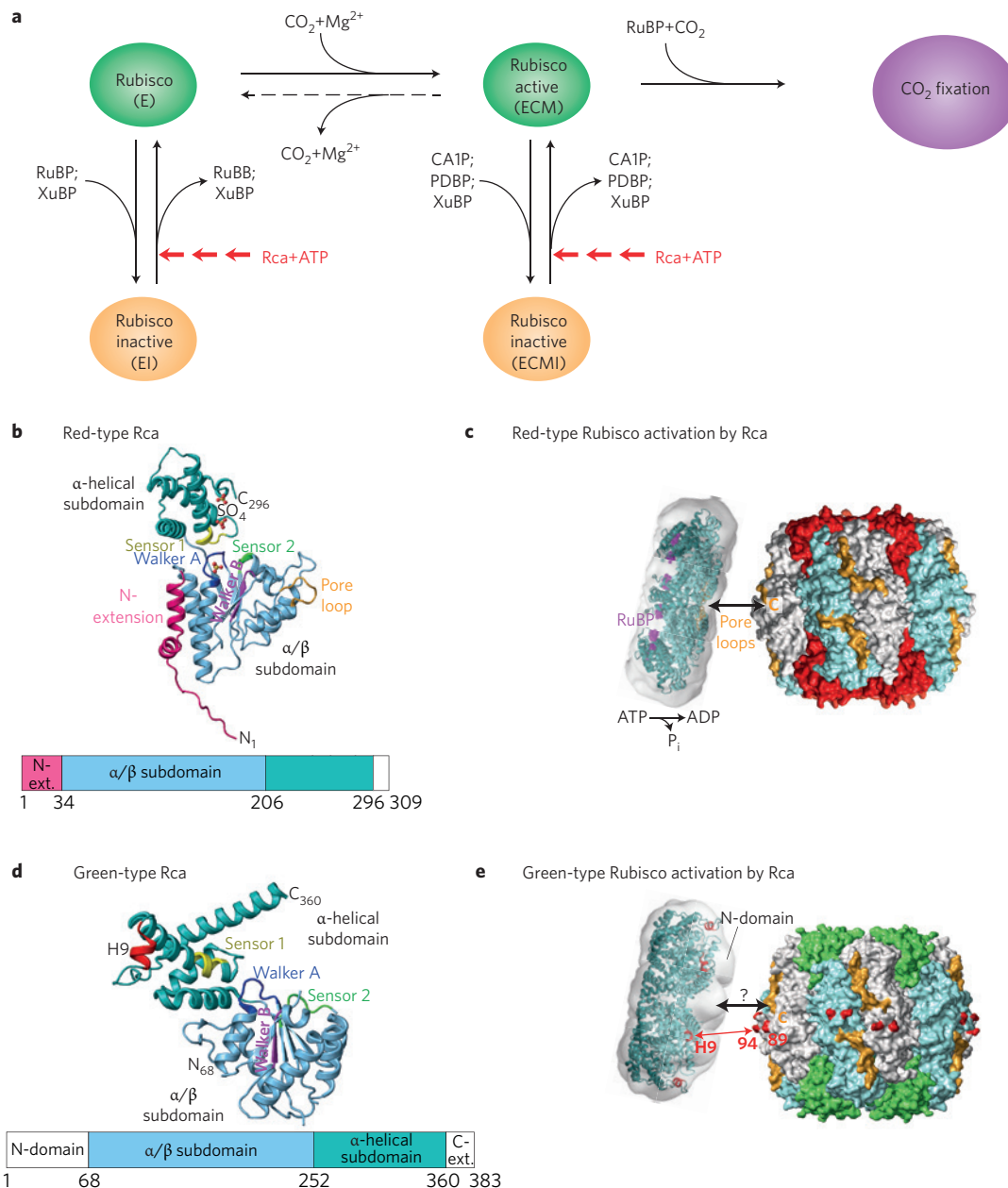


Figure 6 | Rubisco inactivation and reactivation. **a**, Regulation of Rubisco activity and inhibition by sugar phosphates (modified from ref. 68, Oxford Univ. Press). (E), the non-carbamylated enzyme; (ECM), the carbamylated and Mg-bound enzyme; (EI), the inhibited (E) form; (ECMI), the inhibited (ECM) form; PDBP, 2,3-pentodiulose-1,5-bisphosphate. **b**, Crystal structure of the subunit of red-type Rca from the proteobacteria *R. sphaeroides* in ribbon representation (PDB: 3SYK). The α/β and the α -helical subdomains are indicated in blue and teal, respectively, and a N-terminal extension in dark pink. The Walker A and B motifs are shown in dark blue and magenta, the sensor I and II regions in yellow and green, respectively. The pore loop is indicated. Bound sulphates are shown in ball-and-stick. The domain structure is schematically shown as a bar at the bottom. Figure modified from ref. 72, NPG. **c**, Model of the mechanism of red-type Rca. The Rca hexamer docks onto inactive Rubisco and transiently pulls the extended C-terminal sequence of red-type RbcL into the central pore of the hexamer. This action is mediated by the stimulated Rca ATPase and destabilizes the Rubisco active site, releasing the inhibitory sugar. Rca is shown in ribbon representation within the envelope obtained from electron microscopy (EM) reconstruction⁷². The positions of the C-terminal RbcL tails (orange), the Rca pore loops (orange) and the RuBP allosteric sites (magenta) on Rca are indicated. Figure modified from ref. 72, NPG. **d**, Crystal structure of the subunit of green-type Rca from tobacco, lacking the N-domain, in ribbon representation (PDB: 3T15). The α/β and the α -helical subdomains are indicated in blue and teal, respectively. The canonical AAA+ structural motifs are indicated as in **b**. The specificity helix (H9) is shown in red. The domain structure is schematically shown as a bar at the bottom. Figure modified from ref. 73, NPG. **e**, Model of the mechanism of green-type Rca. The interaction of Rca hexamer with inactive Rubisco requires the additional N-domain and the H9 specificity helix (red). Residues in RbcL that are critical for recognition by Rca (ref. 66) are indicated in red. Whether green-type Rca engages with the C-terminal tail of RbcL (orange) is unclear. Rca is shown in ribbon representation within the envelope obtained from EM reconstruction of full-length Rca hexamer⁷³.

Received 15 February 2015; accepted 20 April 2015;
published online 2 June 2015

References

- Andersson, I. & Backlund, A. Structure and function of Rubisco. *Plant Physiol. Biochem.* **46**, 275–291 (2008).
- Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**, 237–240 (1998).
- Miziorko, H. M. & Lorimer, G. H. Ribulose-1,5-bisphosphate carboxylase-oxygenase. *Annu. Rev. Biochem.* **52**, 507–535 (1983).
- Hartman, F. C. & Harpel, M. R. Structure, function, regulation, and assembly of D-Ribulose-1,5-bisphosphate carboxylase/oxygenase. *Annu. Rev. Biochem.* **63**, 197–234 (1994).
- Ellis, R. J. Biochemistry: Tackling unintelligent design. *Nature* **463**, 164–165 (2010).
- Andersson, I. Catalysis and regulation in Rubisco. *J. Exp. Bot.* **59**, 1555–1568 (2008).
- Portis, A. R. & Parry, M. A. J. Discoveries in Rubisco (Ribulose 1,5-bisphosphate carboxylase/oxygenase): a historical perspective. *Photosynth. Res.* **94**, 121–143 (2007).
- Maurino, V. G. & Peterhansel, C. Photorespiration: current status and approaches for metabolic engineering. *Curr. Opin. Plant Biol.* **13**, 249–256 (2010).
- Peterhansel, C., Niessen, M. & Kebeish, R. M. Metabolic engineering towards the enhancement of photosynthesis. *Photochem. Photobiol.* **84**, 1317–1323 (2008).
- Whitney, S. M., Houtz, R. L. & Alonso, H. Advancing our understanding and capacity to engineer nature's CO₂-sequestering enzyme, Rubisco. *Plant Physiol.* **155**, 27–35 (2011).
- Pearce, F. G. Catalytic by-product formation and ligand binding by Ribulose biphosphate carboxylases from different phylogenies. *Biochem. J.* **399**, 525–534 (2006).
- Evans, J. R. & von Caemmerer, S. Enhancing photosynthesis. *Plant Physiol.* **155**, 19 (2011).
- Raines, C. A. Increasing photosynthetic carbon assimilation in C3 plants to improve crop yield: current and future strategies. *Plant Physiol.* **155**, 36–42 (2011).
- Evans, J. R. Improving photosynthesis. *Plant Physiol.* **162**, 1780–1793 (2013).
- Parry, M. A. J. *et al.* Rubisco activity and regulation as targets for crop improvement. *J. Exp. Bot.* **64**, 717–730 (2013).
- Tabita, F. R., Satagopan, S., Hanson, T. E., Kreel, N. E. & Scott, S. S. Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *J. Exp. Bot.* **59**, 1515–1524 (2008).
- Parry, M. A., Andralojc, P. J., Mitchell, R. A., Madgwick, P. J. & Keys, A. J. Manipulation of Rubisco: the amount, activity, function and regulation. *J. Exp. Bot.* **54**, 1321–1333 (2003).
- Spreitzer, R. J. Role of the small subunit in Ribulose-1,5-bisphosphate carboxylase/oxygenase. *Arch. Biochem. Biophys.* **414**, 141–149 (2003).
- Ishikawa, C., Hatanaka, T., Misoo, S., Miyake, C. & Fukayama, H. Functional incorporation of sorghum small subunit increases the catalytic turnover rate of Rubisco in transgenic rice. *Plant Physiol.* **156**, 1603–1611 (2011).
- Tabita, F. R. Microbial Ribulose 1,5-bisphosphate carboxylase/oxygenase: a different perspective. *Photosynth. Res.* **60**, 1–28 (1999).
- Schwenkert, S., Soll, J. & Bolter, B. Protein import into chloroplasts—how chaperones feature into the game. *Biochim. Biophys. Acta* **1808**, 901–911 (2011).
- Whitney, S. M., Baldet, P., Hudson, G. S. & Andrews, T. J. Form I Rubiscos from non-green algae are expressed abundantly but not assembled in tobacco chloroplasts. *Plant J.* **26**, 535–547 (2001).
- Duff, A. P., Andrews, T. J. & Curmi, P. M. The transition between the open and closed states of rubisco is triggered by the inter-phosphate distance of the bound biphosphate. *J. Mol. Biol.* **298**, 903–916 (2000).
- Barracough, R. & Ellis, R. J. Protein synthesis in chloroplasts. IX. Assembly of newly-synthesized large subunits into Ribulose biphosphate carboxylase in isolated intact pea chloroplasts. *Biochim. Biophys. Acta* **608**, 18–31 (1980).
- Hartl, F. U. & Hayer-Hartl, M. The first chaperonin. *Nature Rev. Mol. Cell Biol.* **14**, 611 (2013).
- Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-Hartl, M. & Hartl, F. U. Molecular chaperone functions in protein folding and proteostasis. *Annu. Rev. Biochem.* **82**, 323–355 (2013).
- Brinker, A. *et al.* Dual function of protein confinement in chaperonin-assisted protein folding. *Cell* **107**, 223–233 (2001).
- Liu, C. *et al.* Coupled chaperone action in folding and assembly of hexadecameric Rubisco. *Nature* **463**, 197–202 (2010).
- Vitlin Gruber, A., Nisemlat, S., Azem, A. & Weiss, C. The complexity of chloroplast chaperonins. *Trends Plant Sci.* **18**, 688–694 (2013).
- Weiss, C., Bonshien, A., Farchi-Pisanty, O., Vitlin, A. & Azem, A. Cpn20: siamese twins of the chaperonin world. *Plant Mol. Biol.* **69**, 227–238 (2009).
- Tsai, Y.-C. C., Mueller-Cajar, O., Saschenbrecker, S., Hartl, F. U. & Hayer-Hartl, M. Chaperonin cofactors, Cpn10 and Cpn20, of green algae and plants function as hetero-oligomeric ring complexes. *J. Biol. Chem.* **287**, 20471–20481 (2012).
- Vitlin Gruber, A., Zizelski, G., Azem, A. & Weiss, C. The Cpn10(1) co-chaperonin of *A. thaliana* functions only as a hetero-oligomer with Cpn20. *PLoS ONE* **9**, e113835 (2014).
- Saibil, H. R., Fenton, W. A., Clare, D. K. & Horwich, A. L. Structure and allostery of the chaperonin GroEL. *J. Mol. Biol.* **425**, 1476–1487 (2013).
- Bonshien, A. L. *et al.* Differential effects of co-chaperonin homologs on Cpn60 oligomers. *Cell Stress Chaperones* **14**, 509–519 (2009).
- Kerner, M. J. *et al.* Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell* **122**, 209–220 (2005).
- Goloubinoff, P., Christeller, J. T., Gatenby, A. A. & Lorimer, G. H. Reconstitution of active dimeric Ribulose biphosphate carboxylase from an unfolded state depends on two chaperonin proteins and MgATP. *Nature* **342**, 884–889 (1989).
- Georgescauld, F. *et al.* GroEL/ES chaperonin modulates the mechanism and accelerates the rate of TIM-barrel domain folding. *Cell* **157**, 922–934 (2014).
- Larimer, F. W. & Soper, T. S. Overproduction of *Anabaena* 7120 Ribulose-bisphosphate carboxylase/oxygenase in *Escherichia coli*. *Gene* **126**, 85–92 (1993).
- Emlyn-Jones, D., Woodger, F. J., Price, G. D. & Whitney, S. M. RbcX can function as a Rubisco chaperonin, but is non-essential in *Synechococcus* PCC7942. *Plant Cell Physiol.* **47**, 1630–1640 (2006).
- Andrews, T. J. Catalysis by cyanobacterial Ribulose-bisphosphate carboxylase large subunits in the complete absence of small subunits. *J. Biol. Chem.* **263**, 12213–12219 (1988).
- Saschenbrecker, S. *et al.* Structure and function of RbcX, an assembly chaperone for hexadecameric Rubisco. *Cell* **129**, 1189–1200 (2007).
- Kolesinski, P., Belusiak, I., Czarnocki-Cieciora, M. & Szczepaniak, A. Rubisco accumulation factor 1 from *Thermosynechococcus elongatus* participates in the final stages of Ribulose-1,5-bisphosphate carboxylase/oxygenase assembly in *Escherichia coli* cells and *in vitro*. *FEBS J.* **281**, 3920–3932 (2014).
- Tabita, F. R. Rubisco: The enzyme that keeps on giving. *Cell* **129**, 1039–1040 (2007).
- Li, L.-A. & Tabita, F. R. Maximum activity of recombinant Ribulose 1,5-bisphosphate carboxylase/oxygenase of *Anabaena* sp. strain CA requires the product of the *rbcx* gene. *J. Bact.* **179**, 3793–3796 (1997).
- Onizuka, T. *et al.* The *rbcx* gene product promotes the production and assembly of Ribulose-1, 5-bisphosphate carboxylase/oxygenase of *Synechococcus* sp. PCC7002 in *Escherichia coli*. *Plant Cell Physiol.* **45**, 1390–1395 (2004).
- Tanaka, S., Sawaya, M. R., Kerfeld, C. A. & Yeates, T. O. Structure of the Rubisco chaperone RbcX from *Synechocystis* sp. PCC6803. *Acta Crystallogr. D Biol. Crystallogr.* **63**, 1109–1112 (2007).
- Tarnawski, M., Krzywda, S., Bialek, W., Jaskolski, M. & Szczepaniak, A. Structure of the Rubisco chaperone RbcX from the thermophilic cyanobacterium *Thermosynechococcus elongatus*. *Acta Crystallogr. F Struct. Biol. Cryst. Commun.* **67**, 851–857 (2011).
- Kolesinski, P. *et al.* Insights into eukaryotic Rubisco assembly — Crystal structures of RbcX chaperones from *Arabidopsis thaliana*. *Biochim. Biophys. Acta* **1830**, 2899–2906 (2013).
- Kolesinski, P., Piechota, J. & Szczepaniak, A. Initial characteristics of RbcX proteins from *Arabidopsis thaliana*. *Plant Mol. Biol.* **77**, 447–459 (2011).
- Bracher, A., Starling-Windhof, A., Hartl, F. U. & Hayer-Hartl, M. Crystal structure of a chaperone-bound assembly intermediate of form I Rubisco. *Nature Struct. Mol. Biol.* **18**, 875–880 (2011).
- van Lun, M., van der Spoel, D. & Andersson, I. Subunit interface dynamics in hexadecameric Rubisco. *J. Mol. Biol.* **411**, 1083–1098 (2011).
- Tarnawski, M., Gubernator, B., Kolesinski, P. & Szczepaniak, A. Heterologous expression and initial characterization of recombinant RbcX protein from *Thermosynechococcus elongatus* BP-1 and the role of RbcX in Rubisco assembly. *Acta Biochim. Pol.* **55**, 777–785 (2008).
- Pechmann, S., Willmund, F. & Frydman, J. The ribosome as a hub for protein quality control. *Mol. Cell* **49**, 411–421 (2013).
- Checa, S. K. & Viale, A. M. The 70-Kda Heat-shock protein Dnak chaperone system is required for the productive folding of Ribulose-bisphosphate carboxylase subunits in *Escherichia coli*. *Eur. J. Biochem.* **248**, 848–855 (1997).
- Ivey, R. A. III, Subramanian, C. & Bruce, B. D. Identification of a Hsp70 recognition domain within the rubisco small subunit transit peptide. *Plant Physiol.* **122**, 1289–1299 (2000).

56. Friso, G., Majeran, W., Huang, M., Sun, Q. & van Wijk, K. J. Reconstruction of metabolic pathways, protein expression, and homeostasis machineries across maize bundle sheath and mesophyll chloroplasts: large-scale quantitative proteomics using the first maize genome assembly. *Plant Physiol.* **152**, 1219–1250 (2010).
57. Brutnell, T. P., Sawers, R. J., Mant, A. & Langdale, J. A. Bundle sheath defective2, a novel protein required for post-translational regulation of the *rbcl* gene of maize. *Plant Cell* **11**, 849–864 (1999).
58. Doron, L., Segal, N., Gibori, H. & Shapira, M. The Bsd2 ortholog in *Chlamydomonas reinhardtii* is a polysome-associated chaperone that co-migrates on sucrose gradients with the *rbcl* transcript encoding the Rubisco large subunit. *Plant J.* **80**, 345–355 (2014).
59. Kampinga, H. H. & Craig, E. A. The Hsp70 chaperone machinery: J proteins as drivers of functional specificity. *Nature Rev. Mol. Cell Biol.* **11**, 579–592 (2010).
60. Feiz, L. *et al.* Ribulose-1, 5-bis-phosphate carboxylase/oxygenase accumulation factor1 is required for holoenzyme assembly in maize. *Plant Cell* **24**, 3435–3446 (2012).
61. Wheatley, N. M., Sundberg, C. D., Gidaniyan, S. D., Cascio, D. & Yeates, T. O. Structure and identification of a pterin dehydratase-like protein as a Ribulose-bisphosphate carboxylase/oxygenase (Rubisco) assembly factor in the alpha-carboxysome. *J. Biol. Chem.* **289**, 7973–7981 (2014).
62. Feiz, L. *et al.* A protein with an inactive pterin-4a-carbinolamine dehydratase domain is required for Rubisco biogenesis in plants. *Plant J.* **80**, 862–869 (2014).
63. Whitney, S. M., Birch, R., Kelso, C., Beck, J. L. & Kapralov, M. V. Improving recombinant Rubisco biogenesis, plant photosynthesis and growth by co-expressing its ancillary RAF1 chaperone. *Proc. Natl. Acad. Sci. USA* <http://doi.org/4rj> (2015).
64. Joshi, J., Mueller-Cajar, O., Tsai, Y. C., Hartl, F. U. & Hayer-Hartl, M. Role of small subunit in mediating assembly of red-type form 1 Rubisco. *J. Biol. Chem.* **290**, 1066–1074 (2015).
65. Feller, U., Anders, I. & Mae, T. Rubiscolytics: fate of Rubisco after its enzymatic function in a cell is terminated. *J. Exp. Bot.* **59**, 1615–1624 (2008).
66. Portis, A. R., Li, C. S., Wang, D. F. & Salvucci, M. E. Regulation of Rubisco activase and its interaction with Rubisco. *J. Exp. Bot.* **59**, 1597–1604 (2008).
67. Mueller-Cajar, O., Stotz, M. & Bracher, A. Maintaining photosynthetic CO₂ fixation via protein remodelling: the Rubisco activases. *Photosynth. Res.* **119**, 191–201 (2014).
68. Parry, M. A. J., Keys, A. J., Madgwick, P. J., Carmo-Silva, A. E. & Andralojc, P. J. Rubisco regulation: a role for inhibitors. *J. Exp. Bot.* **59**, 1569–1580 (2008).
69. Hanson, P. I. & Whiteheart, S. W. AAA+ proteins: Have engine, will work. *Nature Rev. Mol. Cell Biol.* **6**, 519–529 (2005).
70. Snider, J., Thibault, G. & Houry, W. A. The AAA+ superfamily of functionally diverse proteins. *Genome Biol.* **9**, 216 (2008).
71. Wendler, P., Ciniawsky, S., Kock, M. & Kube, S. Structure and function of the AAA+ nucleotide binding pocket. *Biochim. Biophys. Acta* **1823**, 2–14 (2012).
72. Mueller-Cajar, O. *et al.* Structure and function of the AAA+ protein CbbX, a red-type Rubisco activase. *Nature* **479**, 194–199 (2011).
73. Stotz, M. *et al.* Structure of green-type Rubisco activase from tobacco. *Nature Struct. Mol. Biol.* **18**, 1366–1370 (2011).
74. Henderson, J. N., Kuriata, A. M., Fromme, R., Salvucci, M. E. & Wachter, R. M. Atomic resolution X-ray structure of the substrate recognition domain of higher plant Ribulose-bisphosphate carboxylase/oxygenase (Rubisco) activase. *J. Biol. Chem.* **286**, 35683–35688 (2011).
75. Bracher, A., Sharma, A., Starling-Windhof, A., Hartl, F. U. & Hayer-Hartl, M. Degradation of potent Rubisco inhibitor by selective sugar phosphatase. *Nature Plants* **1**, 14002 (2015).
76. Linster, C. L., Van Schaftingen, E. & Hanson, A. D. Metabolite damage and its repair or pre-emption. *Nature Chem. Biol.* **9**, 72–80 (2013).
77. Blayney, M. J., Whitney, S. M. & Beck, J. L. NanoESI mass spectrometry of Rubisco and Rubisco activase structures and their interactions with nucleotides and sugar phosphates. *J. Am. Soc. Mass Spectrom.* **22**, 1588–1601 (2011).
78. Chakraborty, M. *et al.* Protein oligomerization monitored by fluorescence fluctuation spectroscopy: self-assembly of rubisco activase. *Biophys. J.* **103**, 949–958 (2012).
79. Keown, J. R., Griffin, M. D. W., Mertens, H. D. T. & Pearce, F. G. Small oligomers of Ribulose-bisphosphate carboxylase/oxygenase (Rubisco) activase are required for biological activity. *J. Biol. Chem.* **288**, 20607–20615 (2013).
80. Henderson, J. N., Hazra, S., Dunkle, A. M., Salvucci, M. E. & Wachter, R. M. Biophysical characterization of higher plant Rubisco activase. *Biochim. Biophys. Acta* **1834**, 87–97 (2013).
81. Kuriata, A. M. *et al.* ATP and magnesium promote cotton short-form Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) activase hexamer formation at low micromolar concentrations. *Biochemistry* **53**, 7232–7246 (2014).
82. Keown, J. R. & Pearce, F. G. Characterization of spinach Ribulose-1,5-bisphosphate carboxylase/oxygenase activase isoforms reveals hexameric assemblies with increased thermal stability. *Biochem. J.* **464**, 413–423 (2014).
83. Wachter, R. M. *et al.* Activation of interspecies-hybrid Rubisco enzymes to assess different models for the Rubisco-Rubisco activase interaction. *Photosynth. Res.* **117**, 557–566 (2013).
84. Sage, R. F., Way, D. A. & Kubien, D. S. Rubisco, Rubisco activase, and global climate change. *J. Exp. Bot.* **59**, 1581–1595 (2008).
85. Nisbet, E. G. *et al.* The age of Rubisco: the evolution of oxygenic photosynthesis. *Geobiology* **5**, 311–335 (2007).
86. Mueller-Cajar, O. & Whitney, S. M. Directing the evolution of Rubisco and Rubisco activase: first impressions of a new tool for photosynthesis research. *Photosynth. Res.* **98**, 667–675 (2008).
87. Maisnier-Patin, S. *et al.* Genomic buffering mitigates the effects of deleterious mutations in bacteria. *Nature Genet.* **37**, 1376–1379 (2005).
88. Tokuriki, N. & Tawfik, D. S. Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* **19**, 596–604 (2009).
89. Williams, T. A. & Fares, M. A. The effect of chaperonin buffering on protein evolution. *Genome Biol. Evol.* **2**, 609–619 (2010).
90. Bogumil, D. & Dagan, T. Cumulative impact of chaperone-mediated folding on genome evolution. *Biochemistry* **51**, 9941–9953 (2012).
91. Wyganowski, K. T., Kaltenbach, M. & Tokuriki, N. GroEL/ES buffering and compensatory mutations promote protein evolution by stabilizing folding intermediates. *J. Mol. Biol.* **425**, 3403–3414 (2013).
92. Tokuriki, N. & Tawfik, D. S. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* **459**, 668–671 (2009).
93. Bershtein, S., Mu, W., Serohijos, A. W. R., Zhou, J. & Shakhnovich, E. I. Protein quality control acts on folding intermediates to shape the effects of mutations on organismal fitness. *Mol. Cell* **49**, 133–144 (2013).
94. Greene, D. N., Whitney, S. M. & Matsumura, I. Artificially evolved *Synechococcus* PCC6301 Rubisco variants exhibit improvements in folding and catalytic efficiency. *Biochem. J.* **404**, 517–524 (2007).
95. Mueller-Cajar, O. & Whitney, S. M. Evolving improved *Synechococcus* Rubisco functional expression in *Escherichia coli*. *Biochem. J.* **414**, 205–214 (2008).
96. Durão, P. *et al.* Opposing effects of folding and assembly chaperones on evolvability of Rubisco. *Nature Chem. Biol.* **11**, 148–155 (2015).
97. Zarzycki, J., Axen, S. D., Kinney, J. N. & Kerfeld, C. A. Cyanobacterial-based approaches to improving photosynthesis in plants. *J. Exp. Bot.* **64**, 787–798 (2013).
98. Lin, M. T., Occhialini, A., Andralojc, P. J., Parry, M. A. & Hanson, M. R. A faster Rubisco with potential to increase photosynthesis in crops. *Nature* **513**, 547–550 (2014).
99. Notredame, C., Higgins, D. G. & Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217 (2000).
100. Huson, D. H. & Scornavacca, C. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Systematic Biol.* **61**, 1061–1067 (2012).

Additional information

Supplementary information is available online. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence should be addressed to M.H.H.

Competing interests

The authors declare no competing financial interests.

1.5 Aim of the study

Rubisco research has drawn much attention recently due to new findings regarding factors that are involved in the assembly process of the enzyme. Until today, no plant Rubisco could be expressed in a functional form in a non-plant host or reconstituted *in vitro*. Recent forecasts suggest that global food production will need to rise by more than 25% by 2050 to meet the ever increasing demand²¹². Engineering a catalytically more efficient Rubisco enzyme could contribute to reaching that goal, but the complex nature of Rubisco's folding and assembly has made these efforts exceedingly challenging. The aim of this study is to structurally and functionally characterize key factors that are involved in the biogenesis of Rubisco. Therefore, the recently discovered putative Rubisco assembly chaperone Raf1 from plant and cyanobacteria was investigated biochemically in its potential role as a Rubisco assembly factor. The role of Raf1 in Rubisco assembly was determined using an *in vitro* reconstitution system. Moreover, the crystal structure of *Arabidopsis thaliana* Raf1 could be solved. Negative stain electron microscopy in combination with chemical cross-linking and mass spectrometric analysis was performed to gain structural insight into the interaction of Raf1 with Rubisco. A comprehensive mutational analysis identified conserved residues of Raf1 that are critical in the assembly process. Finally, native mass spectrometry was used to assess oligomeric states of Raf1 mediated assembly products.

Moreover, eukaryotic RbcX proteins from *Arabidopsis thaliana* and *Chlamydomonas reinhardtii* were biochemically investigated in their ability to assemble form I Rubisco. Furthermore, the crystal structure of *Chlamydomonas reinhardtii* RbcX was solved. Additional mutational analysis revealed a function of eukaryotic RbcX in Rubisco assembly similar to cyanobacterial RbcX. Differences in binding to the C-terminus of the Rubisco large subunit by eukaryotic RbcX compared to cyanobacterial RbcX were observed. Furthermore, first *in vivo* experiments by silencing RbcX in *Chlamydomonas reinhardtii* indicate its potential role in Rubisco assembly.

2. Publications

2.1 Structure and mechanism of the Rubisco-assembly chaperone Raf1

Structure and mechanism of the Rubisco-assembly chaperone Raf1

Thomas Hauser¹, Javaid Y Bhat¹, Goran Miličić¹, Petra Wendler², F Ulrich Hartl¹, Andreas Bracher¹ & Manajit Hayer-Hartl¹

Biogenesis of the photosynthetic enzyme Rubisco, a complex of eight large (RbcL) and eight small (RbcS) subunits, requires assembly chaperones. Here we analyzed the role of Rubisco accumulation factor1 (Raf1), a dimer of ~40-kDa subunits. We find that Raf1 from *Synechococcus elongatus* acts downstream of chaperonin-assisted RbcL folding by stabilizing RbcL antiparallel dimers for assembly into RbcL₈ complexes with four Raf1 dimers bound. Raf1 displacement by RbcS results in holoenzyme formation. Crystal structures show that Raf1 from *Arabidopsis thaliana* consists of a β -sheet dimerization domain and a flexibly linked α -helical domain. Chemical cross-linking and EM reconstruction indicate that the β -domains bind along the equator of each RbcL₂ unit, and the α -helical domains embrace the top and bottom edges of RbcL₂. Raf1 fulfills a role similar to that of the assembly chaperone RbcX, thus suggesting that functionally redundant factors ensure efficient Rubisco biogenesis.

Life on earth depends on the ability of photosynthetic organisms to convert atmospheric CO₂ into organic carbon. The key enzyme responsible for this process is ribulose-1,5-bisphosphate carboxylase-oxygenase (Rubisco). Rubisco catalyzes the rate-limiting step of carbon fixation by carboxylating the five-carbon sugar substrate ribulose-1,5-bisphosphate (RuBP). In plants, algae, cyanobacteria and some proteobacteria, Rubisco (form I) is an oligomeric complex of eight large (RbcL, ~50 kDa) and eight small (RbcS, ~15 kDa) subunits¹. The RbcL subunits are arranged as a tetramer of antiparallel dimers (RbcL₈ core) capped by four RbcS subunits at the top and four at the bottom (RbcL₈S₈ holoenzyme). Remarkably, Rubisco is a rather inefficient enzyme, fixing only around three to ten carbon atoms per second^{2,3}. Moreover, certain Rubiscos confuse CO₂ with molecular oxygen as often as once in four catalytic cycles⁴, thus resulting in wasteful photorespiration^{5–7}. Hence, engineering a more efficient Rubisco enzyme is central to efforts to increase global food production^{8–11}.

Although the structure and mechanism of Rubisco are well understood^{12,13}, knowledge of the chaperone machinery required for Rubisco biogenesis remains limited. As recently demonstrated by *in vitro* reconstitution¹⁴, folding of cyanobacterial RbcL requires the chaperonin GroEL and its cofactor GroES (Cpn60, Cpn10 and Cpn20 in chloroplasts¹⁵). The spontaneous assembly of folded RbcL is inefficient and has been shown to be facilitated by a specific assembly chaperone, RbcX^{14,16}. In several cyanobacteria, RbcX is encoded in an operon between the *rbcL* and *rbcS* genes, and coexpression with RbcX has been shown to enhance the production of active form I Rubisco in *Escherichia coli*^{16–21}. RbcX, a homodimer of ~15-kDa subunits, functions as a molecular clamp in stabilizing the antiparallel RbcL dimer and promotes formation of an RbcL₈–RbcX₈ assembly

intermediate^{14,22} from which RbcX is displaced by RbcS¹⁴. However, RbcX is not essential in *S. elongatus* PCC7942 (Syn7942)²³, thus suggesting functional redundancy with other factors. Indeed, a recent screen of photosynthetic mutants in maize identified a nuclear-encoded chloroplast protein, Raf1 (Rubisco accumulation factor 1), that is required for efficient Rubisco biogenesis²⁴. Raf1 is conserved in all photosynthetic organisms containing RbcX and functions in Rubisco assembly *in vitro* and *in vivo*^{25,26}.

Here we set out to functionally and structurally characterize the plant and cyanobacterial Raf1 proteins. We solved the crystal structures of the *A. thaliana* Raf1 domains and analyzed the interaction of Raf1 with RbcL by multiple biochemical and biophysical approaches. Our results show that Raf1 brackets the antiparallel RbcL dimer and stabilizes it in a state competent for assembly of higher oligomers up to RbcL₈. Displacement of Raf1 by RbcS leads to formation of the functional holoenzyme. Thus, Raf1 fulfills a role similar to that of RbcX but uses a different structural scaffold and mechanism.

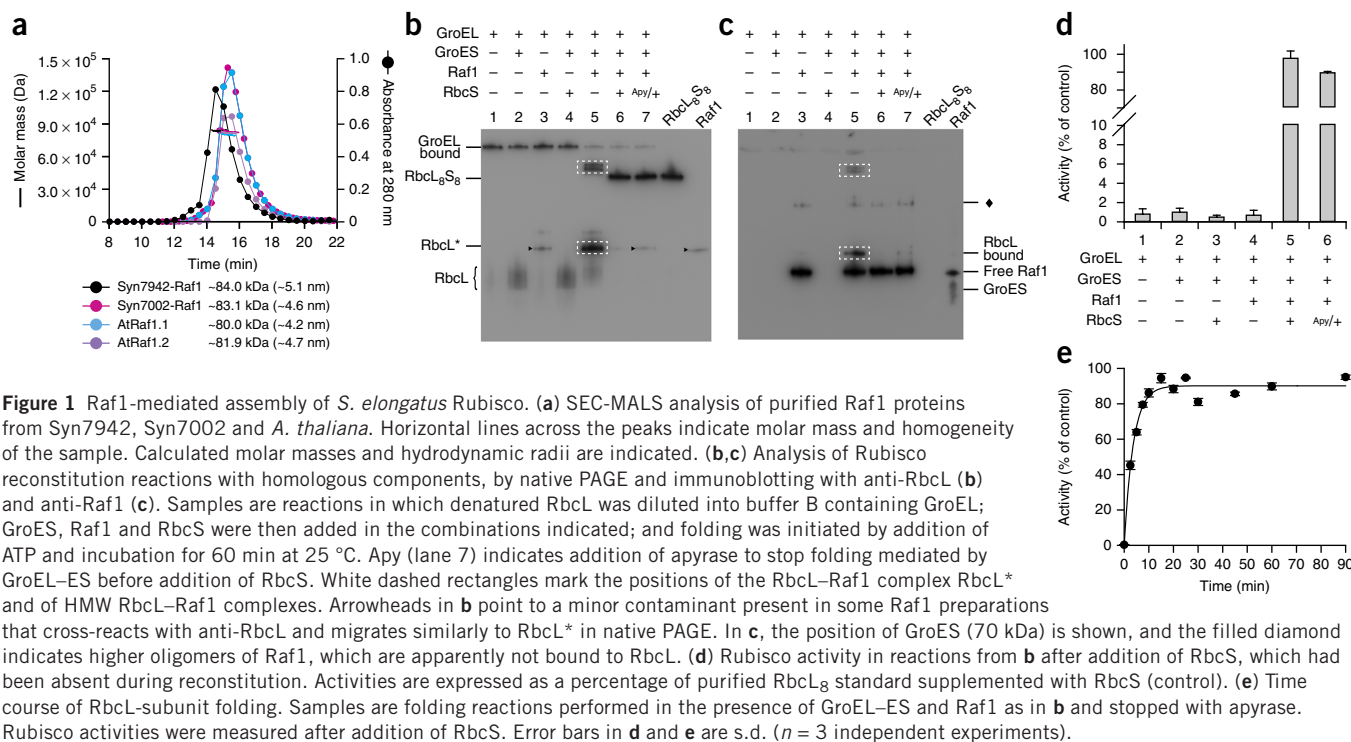
RESULTS

Reconstitution of Raf1 function in Rubisco assembly

A. thaliana (At) contains two Raf1 homologs encoded by the *AT3G04550* (*Atraf1.2*) and *AT5G28500* (*Atraf1.1*) genes, whereas cyanobacteria have only one *raf1* gene. The respective Raf1 proteins share ~38% sequence similarity, and their cyanobacterial homologs are ~55% similar (**Supplementary Fig. 1**). We recombinantly expressed AtRaf1.1 (42.6 kDa) and AtRaf1.2 (43.5 kDa) without their predicted transit peptides, as well as Raf1 from *S. elongatus* PCC7942 (Syn7942-Raf1; 40.2 kDa) and *Synechococcus* sp. PCC7002 (Syn7002-Raf1; 39.9 kDa). The purified Raf1 proteins (**Supplementary Fig. 2a**) formed homodimers, as determined by size-exclusion chromatography

¹Department of Cellular Biochemistry, Max Planck Institute of Biochemistry, Martinsried, Germany. ²Gene Center Munich, Ludwig-Maximilians-Universität München, Munich, Germany. Correspondence should be addressed to A.B. (bracher@biochem.mpg.de) or M.H.-H. (mhartl@biochem.mpg.de).

Received 22 October 2014; accepted 1 July 2015; published online 3 August 2015; doi:10.1038/nsmb.3062



combined with static light scattering (SEC-MALS) (Fig. 1a). Henceforth, concentrations of Raf1 refer to the dimer.

To analyze the function of Raf1, we performed reconstitution experiments with RbcL, RbcS and Raf1 from *S. elongatus* PCC7942 (Syn7942). Upon folding of denatured RbcL with GroEL–GroES (GroEL–ES) in the presence of ATP, most RbcL remained associated with GroEL, as analyzed by native PAGE (Fig. 1b, lanes 1 and 2). Only a small amount of RbcL was released from GroEL, and it migrated as a diffuse band (Fig. 1b, lane 2). This species failed to assemble into holoenzyme with RbcS (Fig. 1b, lane 4). Addition of Raf1 at equimolar concentration to RbcL resulted in efficient release of RbcL from GroEL–ES and in formation of a well-defined RbcL species (denoted RbcL*) (Fig. 1b, lane 5). In addition, we detected small amounts of high-molecular-weight (HMW) complexes migrating above the holoenzyme standard (RbcL₈S₈) (Fig. 1b, lane 5). Both RbcL* and the HMW forms represented complexes of RbcL with Raf1, as detected by immunoblotting with anti-Raf1 (in which antibody reactivity with Raf1 was reduced in the HMW complexes) (Fig. 1c, lane 5). The RbcL–Raf1 complexes formed only in the presence of GroEL and GroES (Fig. 1b, lanes 3 and 5), thus indicating that Raf1 acts on folded RbcL. Efficient formation of enzymatically active RbcL₈S₈ occurred in the presence of Raf1 and RbcS (Fig. 1b, lane 6, and Fig. 1d, lane 5). We also observed holoenzyme assembly when RbcS was added after treatment with apyrase, which hydrolyzes ATP to AMP and stops RbcL folding by GroEL–ES (Fig. 1b, lane 7, and Fig. 1d, lane 6). Thus, Raf1 maintained folded RbcL in an assembly-competent state. RbcL folding occurred with an apparent half-time of ~3 min (Fig. 1e), a result consistent with previous measurements¹⁴.

Assembly-competent RbcL* also formed during reconstitution with the heterologous Syn7002-Raf1 or AtRaf1.2 but not with AtRaf1.1 (Supplementary Fig. 2b). Reconstitution was ~70% efficient with Syn7002-Raf1 and only ~10% efficient with AtRaf1.2. We obtained no active enzyme with AtRaf1.1 or a heterodimer of AtRaf1.1 and

AtRaf1.2 (Supplementary Fig. 2c). Thus, Raf1 proteins appear to be adapted to their cognate RbcL proteins, results consistent with those in a recent report²⁶.

RbcL–Raf1 assembly intermediates

The results described above showed that Raf1 assists in the formation of RbcL assembly intermediates that are competent to form holoenzyme with RbcS. Next we tested whether Raf1 could also bind to preformed RbcL₈. Recombinantly produced RbcL₈ migrated as a stable complex in native PAGE¹⁶ (Fig. 2a, lane 1). Addition of Raf1 resulted in an upshift of RbcL₈ indicative of the formation of RbcL₈–Raf1 complexes (Fig. 2a, lane 3), which are similar to the HMW complexes obtained during reconstitution (Fig. 2a, lane 2). RbcS displaced Raf1, thus resulting in holoenzyme formation (Supplementary Fig. 2d). Binding of Raf1 to RbcL₈ also produced a small amount of RbcL–Raf1 complex that migrated indistinguishably from RbcL* (Fig. 2a,b, lanes 3). Raf1 did not interact with RbcL₈S₈ (Fig. 2a,b, lanes 6), and RbcL₈ remained stable in the presence of GroEL (Fig. 2a,b, lanes 7).

To determine the relative stoichiometry of RbcL and Raf1 in the RbcL* complex, we excised RbcL* from native PAGE gels and reanalyzed it by SDS-PAGE, which was followed by Coomassie staining and densitometry. Taking into account that RbcL stains approximately two-fold more intensely than Raf1, this analysis suggests that RbcL* is a complex of one Raf1 dimer and two RbcL subunits (Supplementary Fig. 2e), consistently with the antiparallel RbcL dimer being the building block of RbcL₈.

Next we characterized the HMW complex consisting of RbcL and Raf1 by SEC-MALS and native mass spectrometry (native MS). The molar mass of RbcL₈ was ~390 kDa by SEC-MALS and ~420 kDa by native MS, values close to the theoretical mass of 420.4 kDa (Fig. 2c,d). Addition of Raf1 to RbcL₈ resulted in a complex of ~764 kDa by SEC-MALS and ~741 kDa by native MS (Fig. 2c,d), consistently with four Raf1 dimers being bound to RbcL₈ (theoretical mass 740.9 kDa). SEC-MALS also detected a species of ~159 kDa,

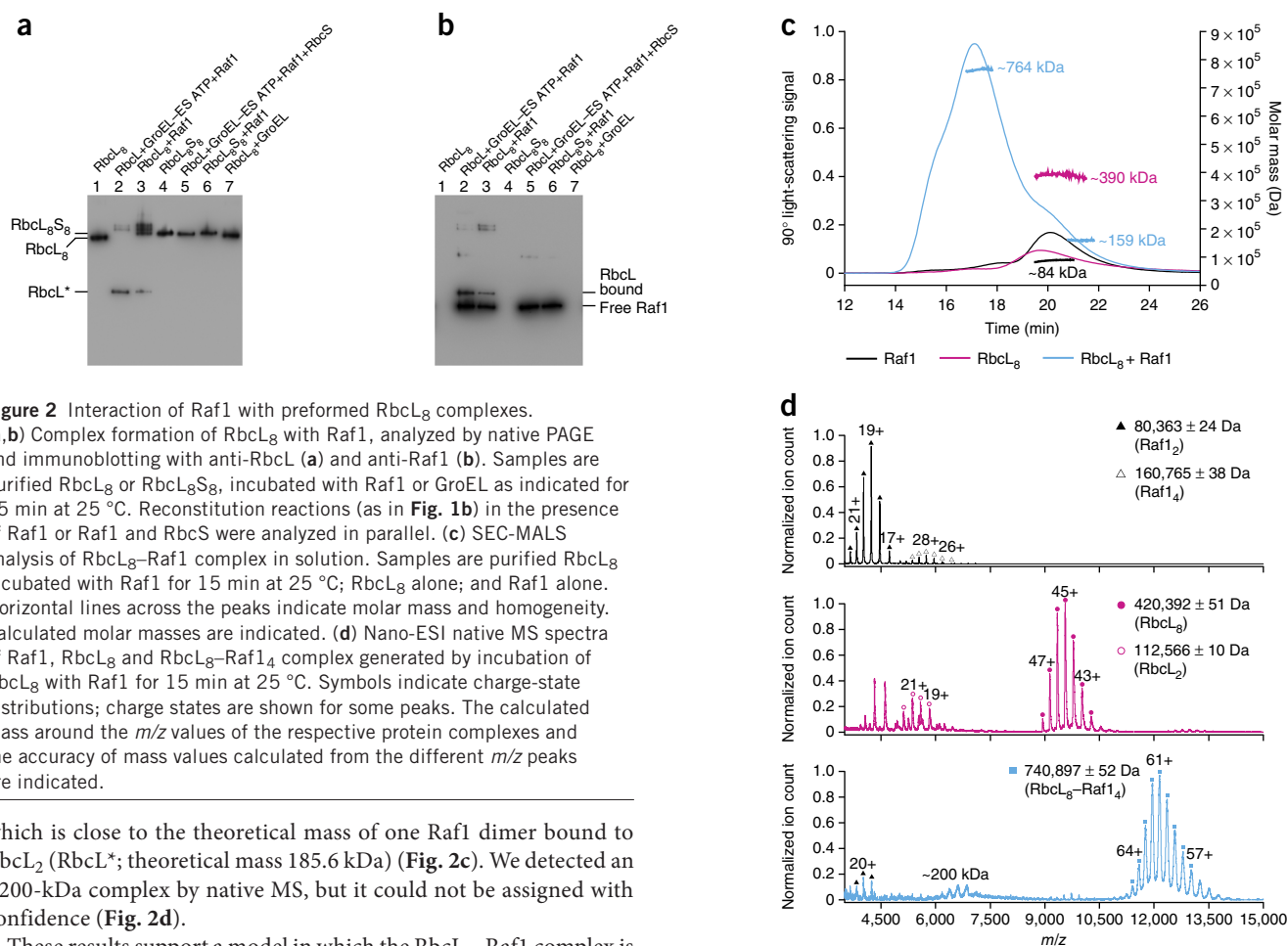


Figure 2 Interaction of Raf1 with preformed RbcL₈ complexes. (a,b) Complex formation of RbcL₈ with Raf1, analyzed by native PAGE and immunoblotting with anti-RbcL (a) and anti-Raf1 (b). Samples are purified RbcL₈ or RbcL₈S₈, incubated with Raf1 or GroEL as indicated for 15 min at 25 °C. Reconstitution reactions (as in Fig. 1b) in the presence of Raf1 or Raf1 and RbcS were analyzed in parallel. (c) SEC-MALS analysis of RbcL₈-Raf1 complex in solution. Samples are purified RbcL₈ incubated with Raf1 for 15 min at 25 °C; RbcL₈ alone; and Raf1 alone. Horizontal lines across the peaks indicate molar mass and homogeneity. Calculated molar masses are indicated. (d) Nano-ESI native MS spectra of Raf1, RbcL₈ and RbcL₈-Raf1₄ complex generated by incubation of RbcL₈ with Raf1 for 15 min at 25 °C. Symbols indicate charge-state distributions; charge states are shown for some peaks. The calculated mass around the m/z values of the respective protein complexes and the accuracy of mass values calculated from the different m/z peaks are indicated.

which is close to the theoretical mass of one Raf1 dimer bound to RbcL₂ (RbcL*; theoretical mass 185.6 kDa) (Fig. 2c). We detected an ~200-kDa complex by native MS, but it could not be assigned with confidence (Fig. 2d).

These results support a model in which the RbcL₂-Raf1 complex is the first assembly intermediate, and the RbcL₈-Raf1₄ complex is the end state of Raf1-mediated assembly. However, the latter was only a minor species during reconstitution in the absence of RbcS, and the equilibrium was shifted to the RbcL₂-Raf1 complex (Fig. 2a,b, lane 2). Notably, the yield of active Rubisco decreased when Raf1 was present in excess over RbcL (Supplementary Fig. 2f).

Crystal structures of Raf1 domains

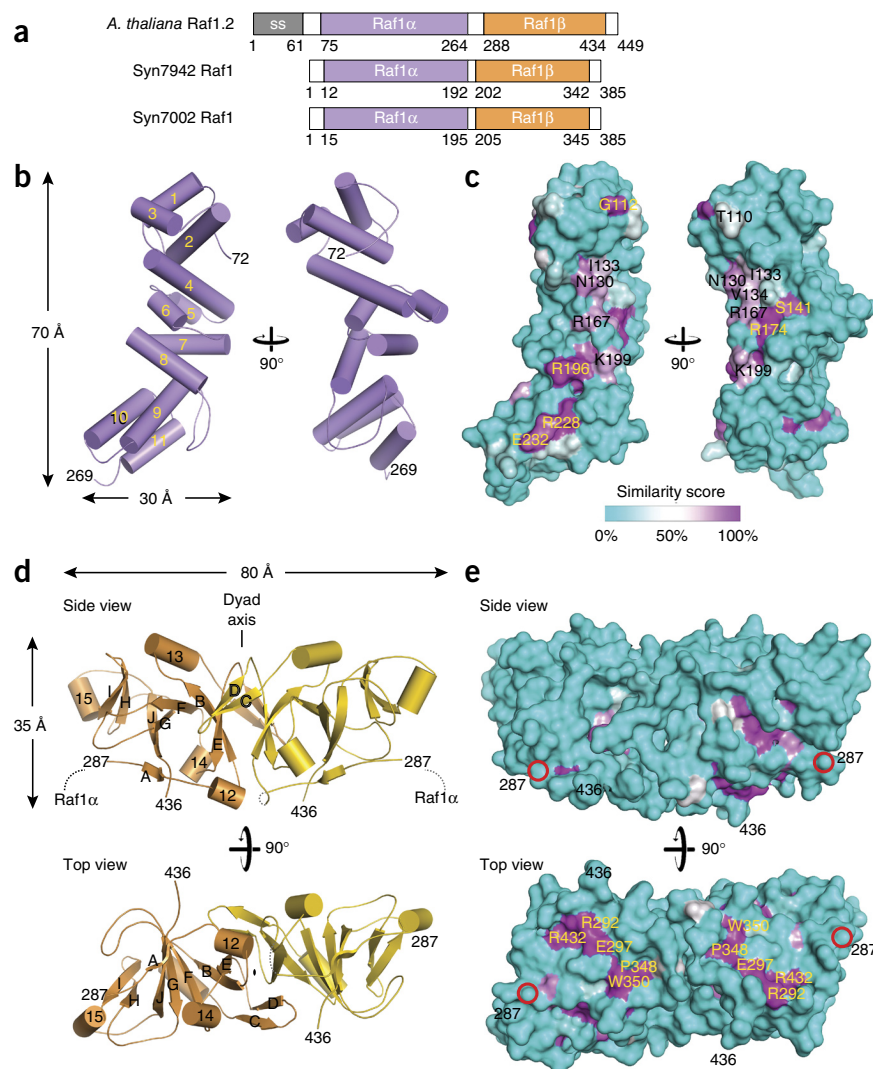
Secondary-structure prediction suggested that Raf1 proteins consist of an α -helical N-terminal domain of ~185 residues and a C-terminal β -sheet domain of ~150 residues, connected by a linker of ~10–27 residues (Fig. 3a). Although Syn7942-Raf1 failed to crystallize, we obtained crystals of AtRaf1.2 containing either residues 62–275 (Raf1 α) or 281–449 (Raf1 β), as determined by MS analysis, thus indicating protease sensitivity of the interdomain linker. We cloned and recombinantly expressed the respective α - and β -domains of the different Raf1 homologs for further analysis (Supplementary Fig. 2a). The Raf1 α domains were monomeric, and the Raf1 β domains behaved as dimers (Supplementary Fig. 3a). The domains on their own were essentially inactive in the Rubisco reconstitution assay (Supplementary Fig. 4).

We solved the crystal structures of AtRaf1.2 α and AtRaf1.2 β by single isomorphous replacement with anomalous scattering (SIRAS) at resolutions of 1.95 Å and 2.6–2.8 Å, respectively (Table 1 and Supplementary Fig. 5a,b). Raf1 α is a slightly curved rod with dimensions of 30 × 30 × 70 Å, composed of 11 stacked α -helices, the last three of which form a regular three-helix bundle (Fig. 3b). Only one face of the domain exhibits substantial surface-residue conservation

among Raf1 homologs; it forms a rather flat surface with hydrophobic and positively charged regions (Fig. 3c and Supplementary Fig. 5c). Particularly conspicuous is the hydrophobic face of helix 4 (Asn130, Ile133, Val134 and Ser141) and the conserved charged residues (Arg167, Arg174, Arg196, Lys199, Arg228 and Glu232) of the adjacent helices 8 and 9.

We obtained two crystal forms for AtRaf1.2 β of space groups C2 (2.8-Å resolution) and P2₁2₁2₁ (2.6-Å resolution) (Table 1). In both crystal forms, Raf1 β exhibited mostly β -structured dimers with pseudo-two-fold symmetry (Fig. 3d). The three independent copies were closely similar (r.m.s. deviation 0.469–1.139 Å), except for the domain swapping of a loop in the P2₁2₁2₁ crystal form (Supplementary Fig. 5d,e). The overall appearance of the dimer was again rod shaped, with dimensions of 30 × 35 × 80 Å, and with the monomer core forming a curved, mixed β -sheet composed of strands A–J–G–F–B (Fig. 3d, side view). Satellite β -sheets (strands I–H–J and B and E) branch off the elongated β -strands B and J and curl back onto the central sheet on both sides. Three short α -helices are interspersed between the strands. The two-fold symmetry is broken where the connectors between the first and second β -strand meet, close to the dyad axis (Fig. 3d and Supplementary Fig. 5f). The dimer interface is formed by the β -hairpin protrusions (strands C and D) from one subunit and β -strands B and E from the other. The interface is highly hydrophobic, and it buries 1,400 Å² in each subunit (Supplementary Fig. 5g). Only the face of the dimer from which the linkers to the α -domains extend is conserved among Raf1 homologs (Fig. 3e). This surface is rather flat, and it contains

Figure 3 Crystal structures of Raf1 domains. (a) Schematic representation of the domain structures of Raf1 from *A. thaliana* (isoform 2; AtRaf1.2), *S. elongatus* PCC7942 (Syn7942) and *Synechococcus* sp. PCC7002 (Syn7002). The predicted chloroplast signal sequence (SS) of AtRaf1.2 is shown in gray. The α -helical and β -sheet domains are shown purple and orange, respectively. Variable linker regions are indicated in white. The domain boundaries for the cyanobacterial Raf1 proteins are based on the sequence alignment in **Supplementary Figure 1**. (b) Crystal structure of AtRaf1.2 α . Views related by 90° rotation are shown. Helices are represented as cylinders. (c) Surface conservation in AtRaf1.2 α . AtRaf1.2 α is oriented as in b. (d) Structure of the AtRaf1.2 β dimer. The two subunits are shown in orange and yellow. Secondary-structure elements are indicated by numbers for α -helices and by letters for β -strands. The position of the pseudo-two-fold axis is shown. (e) Surface conservation in the AtRaf1.2 β dimer, analyzed as in c.



conserved hydrophobic (Pro348 and Trp350) and charged (Arg292, Glu297 and Arg432) residues in adjacent β -strands A and J. The C-terminal 12 residues were disordered in all crystal lattices.

Small-angle X-ray scattering (SAXS) measurements indicated maximum dimensions (D_{\max}) of the isolated domains similar to those derived from the crystal structures (**Supplementary Fig. 3b–d**). The D_{\max} of the full-length AtRaf1.2 dimer was ~ 208 Å, and the radius of gyration (R_g) was 52 Å; we obtained similar values for Syn7942-Raf1. This suggested that the highly charged, flexible linker allows dynamic movements of the Raf1 α domains relative to the β -domain dimer, as supported by an ensemble model of the AtRaf1.2 structure (**Supplementary Fig. 3e**).

Mutational analysis of Raf1

Next we performed a mutational analysis to determine the relevance of specific Raf1 residues for interaction with RbcL. Relative to wild-type Raf1, all mutants of the α -domain, except E159A, displayed enhanced formation of HMW RbcL₈-Raf1 complexes at the expense of the RbcL₂-Raf1 intermediate (**Fig. 4a**). In contrast, the β -domain mutants almost exclusively populated the RbcL₂-Raf1 complex, similarly to wild-type (**Fig. 4a**). Although most of the Raf1 mutants supported holoenzyme assembly upon addition of RbcS, two mutants with conserved positive charges in the α -domain (R104Q and K126A K129A, equivalent to Arg174, Arg196 and Lys199 in AtRaf1.2), resulted in a 45–70% reduced yield of activity (**Fig. 4b,c**). Mutant R104Q showed lower amounts of HMW RbcL-Raf1 complex and an increase in diffusely migrating, low-molecular-weight RbcL, presumably representing unassembled protein (**Fig. 4b**). This suggested that decreased binding affinity of Raf1 R104Q for RbcL generated unstable assembly intermediates. We also noted that several of the β -domain mutants resulted in incomplete conversion of RbcL₂-Raf1 to RbcL₈S₈ (**Fig. 4b**), which correlated with a milder reduction in the yield of active enzyme by 10–20% (**Fig. 4c**).

When taken together, our results suggest that mutations in the α -domain have a more pronounced effect on the binding properties of Raf1 than mutations in the β -domain, consistently with the larger conserved surface of the α -domains. The strength of the RbcL-Raf1 interaction appears to be carefully tuned to allow efficient holoenzyme formation.

Analysis of RbcL-Raf1 interaction by chemical cross-linking

To identify the contact regions between Raf1 and RbcL, we next performed cross-linking coupled to mass spectrometry (CXMS). We added Raf1 to *S. elongatus* RbcL₈ to form the RbcL₈-Raf1₄ complex and then incubated the complex with a 1:1 H₁₂ and D₁₂ isotopic mixture of the lysine-specific cross-linker disuccinimidylsuberate (DSS) (**Supplementary Fig. 6**). We performed these experiments either with the homologous Syn7942-Raf1 or with the functionally active heterologous Syn7002-Raf1 (**Supplementary Fig. 2b,c**), which differ considerably in the number and distribution of lysine residues (**Supplementary Fig. 1**). We separated cross-linked products by SDS-PAGE and analyzed bands >170 kDa, which were likely to contain both Raf1 and RbcL (**Supplementary Fig. 6d**). The median C α -C α distance spanned by DSS is ~ 16.4 Å, with an upper boundary of ~ 36 Å for the structurally most dynamic regions²⁷.

In total, we identified 39 cross-linked peptide pairs in the presence of Syn7942-Raf1 and 45 in the presence of Syn7002-Raf1

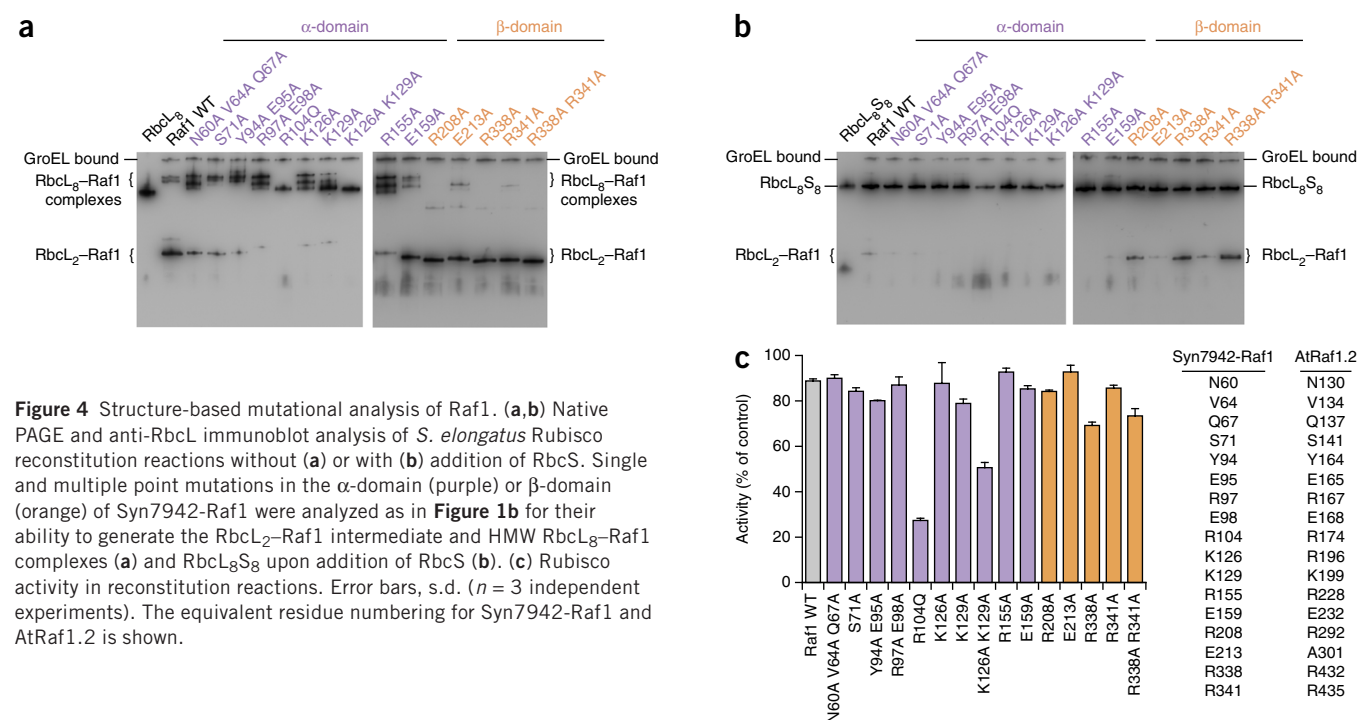
Table 1 Data collection and refinement statistics

| | AtRaf1.2 α native | AtRaf1.2 α K ₂ [PtCl ₄] | AtRaf1.2 β native I | AtRaf1.2 β HMBA-I | AtRaf1.2 β native II |
|---|--|---|---------------------------|-------------------------|---|
| Data collection | | | | | |
| Space group | <i>P</i> 4 ₁ 2 ₁ 2 | <i>P</i> 4 ₁ 2 ₁ 2 | <i>C</i> 2 | <i>C</i> 2 | <i>P</i> 2 ₁ 2 ₁ 2 ₁ |
| Cell dimensions | | | | | |
| <i>a</i> , <i>b</i> , <i>c</i> (Å) | 29.77, 29.77, 457.09 | 29.36, 29.36, 454.86 | 157.54, 34.36, 106.89 | 155.72, 34.70, 115.86 | 39.68, 60.79, 143.27 |
| α , β , γ (°) | 90, 90, 90 | 90, 90, 90 | 90, 93.67, 90 | 90, 97.84, 90 | 90, 90, 90 |
| | | <i>Peak</i> | | <i>Peak</i> | |
| Wavelength | | 1.072 | | 1.009 | |
| Resolution (Å) ^a | 45.71–1.95 (2.06–1.95) ^a | 45.49–2.9 (3.05–2.9) | 45.51–2.8 (2.96–2.8) | 49.38–3.4 (3.58–3.4) | 47.76–2.57 (2.71–2.57) |
| <i>R</i> _{merge} | 0.068 (1.067) | 0.158 (0.706) | 0.044 (0.697) | 0.102 (0.941) | 0.097 (0.909) |
| <i>I</i> / σ <i>I</i> | 20.2 (2.1) | 10.9 (2.9) | 20.0 (1.9) | 19.4 (2.6) | 11.9 (2.1) |
| Completeness (%) | 100 (99.8) | 99.7 (98.2) | 99.2 (96.5) | 99.7 (98.3) | 96.7 (81.8) |
| Redundancy | 13.3 (12.6) | 14.4 (12.9) | 3.6 (3.5) | 11.0 (11.2) | 4.3 (4.2) |
| Refinement | | | | | |
| Resolution (Å) | 30–1.95 | – | 30–2.8 | – | 30–2.57 |
| No. reflections | 16,514 | – | 13,676 | – | 10,713 |
| <i>R</i> _{work} / <i>R</i> _{free} | 0.211 / 0.246 | – | 0.240 / 0.289 | – | 0.210 / 0.279 |
| No. atoms | | | | | |
| Protein | 1,630 | – | 4,443 | – | 2,282 |
| Phosphates ^b | – | – | 15 | – | – |
| Water | 60 | – | – | – | 9 |
| <i>B</i> factors | | | | | |
| Protein | 64.52 | – | 94.61 | – | 60.44 |
| Phosphates | – | – | 138.06 | – | – |
| Water | 51.42 | – | – | – | 45.77 |
| r.m.s. deviations | | | | | |
| Bond lengths (Å) | 0.011 | – | 0.004 | – | 0.008 |
| Bond angles (°) | 1.210 | – | 1.063 | – | 1.272 |

^aValues in parentheses are for highest-resolution shell. ^bFrom precipitant.

(Supplementary Data Set 1a–d and Supplementary Note). We were able to assign plausible intra- or intermolecular distances of 7 to 20 Å for 12 out of the 29 RbcL–RbcL cross-links, on the basis of the

RbcL₈S₈ crystal structure²⁸. The remaining 17 RbcL–RbcL cross-links could not be assigned distances because one or both of the cross-linked residues are located at the flexible N or C terminus. We identified



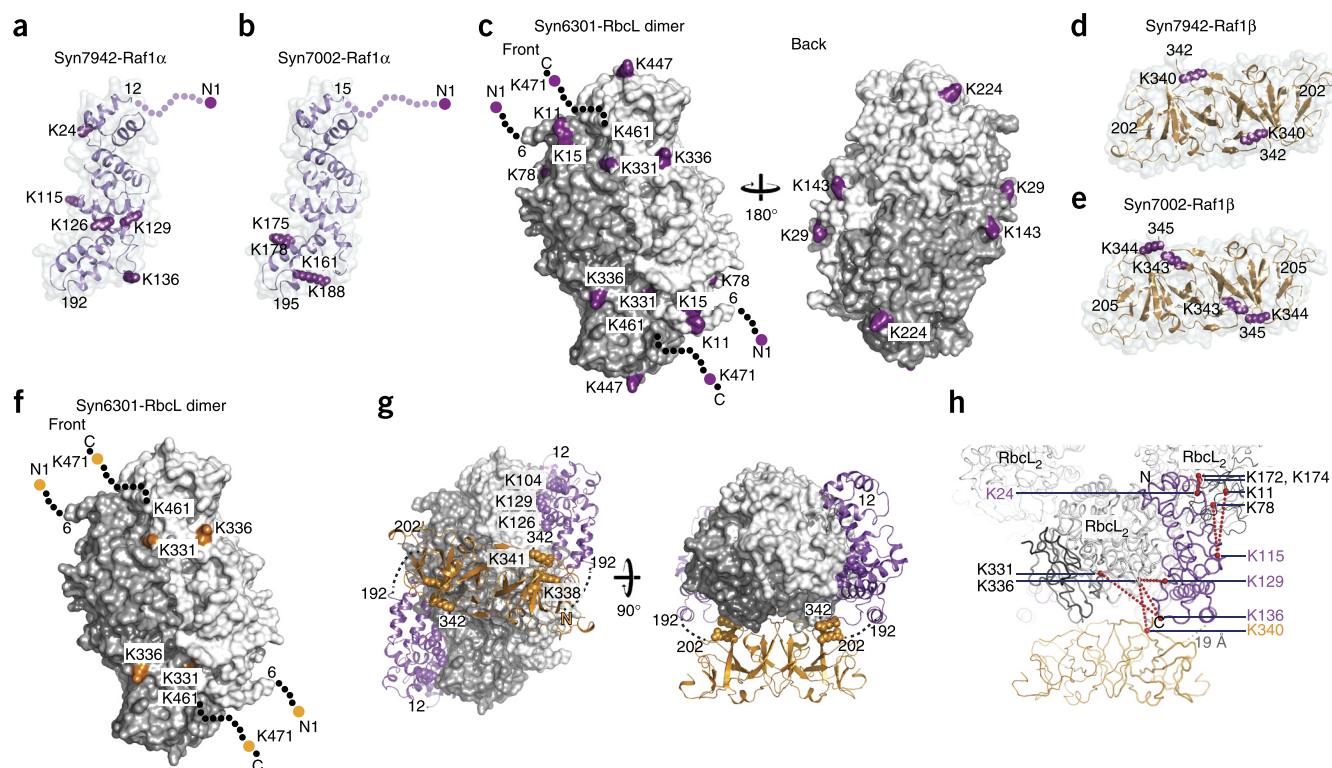


Figure 5 Probing the RbcL–Raf1 complex by chemical cross-linking. **(a,b)** Cross-linking sites in the Raf1 α -domains of Syn7942-Raf1 and Syn7002-Raf1. Cross-linked lysine residues and the N-terminal amino group are shown in space-filling representation. The backbone is shown in ribbon representation underneath a translucent molecular surface. Disordered residues are indicated by dots. The orientation corresponds to **Figure 3b**, left. **(c)** Cross-linking sites from Raf1 α on the surface of the RbcL dimer. The model is based on the crystal structure of Rubisco²⁸. The RbcL subunits are shown in white and gray. Residues Lys172 and Lys174 of RbcL cross-linked to Raf1 α are not indicated because they are partially buried in the RbcL dimer interface, and Lys161 is pointing inwards. **(d,e)** Cross-linking sites in the Raf1 β domains of Syn7942-Raf1 and Syn7002-Raf1, displayed in the same style as in **a** and **b**. Top views are shown. **(f)** Cross-linking sites from Raf1 β on the surface of the RbcL dimer. **(g)** Tentative model of the RbcL₂–Raf1 protomer of the RbcL₈–Raf1₄ complex, based on the cross-linking data. The RbcL dimer is depicted as above; Raf1 is in ribbon representation. Functionally critical Raf1 residues are shown in space-filling representation. **(h)** Cross-links mapped onto the RbcL₈–Raf1₄ complex. Dotted red lines indicate plausible cross-links between lysine residues of RbcL subunits and Raf1.

31 cross-links between RbcL and Raf1, 19 involving the α -domain, 6 involving the β -domain, 4 involving the flexible N terminus and 2 involving the interdomain linker of Raf1 (**Supplementary Data Set 1b,d** and **Supplementary Note**).

According to homology models for the Syn7942-Raf1 and Syn7002-Raf1 domains, most of the α -domain cross-links originated from the periphery of the conserved surface, and we found them multiple times (**Fig. 3c**, **Supplementary Data Set 1b,d** and **Supplementary Note**). They connected to the N domain of RbcL, including the flexible N terminus and residues Lys11, Lys15, Lys29, Lys78 and Lys143, as well as to the TIM-barrel domain (Lys161, Lys172, Lys174, Lys224, Lys331 and Lys336) and the C-terminal domain (Lys447 and Lys471) (**Fig. 5a–c**, **Supplementary Data Set 1b,d** and **Supplementary Note**). The cross-links from the β -domain also involved lysines near its conserved top surface (**Figs. 3e** and **5d,e**). The Syn7942-Raf1 β domain has only one lysine (Lys340), which cross-linked repeatedly to Lys336 on RbcL (**Fig. 5d,f**, **Supplementary Data Set 1b** and **Supplementary Note**). The β -domain of Syn7002-Raf1 has six lysines, and of these Lys343 cross-linked exclusively to Lys336 on RbcL, whereas Lys344 cross-linked to either Lys331 or Lys336 on RbcL (**Fig. 5e,f** and **Supplementary Data Set 1d** and **Supplementary Note**). Both the α - and β -domains cross-linked to residues Lys331 and Lys336 of RbcL (**Supplementary Data Set 1b,d** and **Supplementary Note**), thus suggesting that the Raf1 domains are in proximity, in accordance with the limited linker length of ten residues.

Structural model of the RbcL–Raf1 interaction

We obtained plausible distances for the cross-links between the β -domain and RbcL (19–21 Å) by positioning the β -domain dimer coaxially at the equatorial front face of each RbcL antiparallel dimer (**Fig. 5g,h**, **Supplementary Data Set 1b,d** and **Supplementary Note**). In contrast, positioning the β -domain dimer on the two-fold axis between RbcL dimers resulted in substantially increased cross-link distances (43–67 Å; data not shown), and thus such a topology seems less plausible. Indeed, we found similar cross-links during folding and assembly (**Supplementary Data Set 1e,f** and **Supplementary Note**), where mostly the RbcL₂–Raf1 intermediate was populated (**Fig. 2a,b**). From the equatorial positioning of the β -domains, and taking the length of the α – β domain linker into account, we obtained optimal distances for the cross-links between the α -domain and RbcL (**Supplementary Data Set 1b,d** and **Supplementary Note**) by placing the α -domains so that they embraced the top and bottom edges of each RbcL₂ unit. The shallow groove in the conserved surface of the α -domain (**Fig. 3c**) would cradle the back of the C-terminal domain of RbcL, consistently with mutations in this region weakening the interaction with RbcL (**Fig. 5g** and **Fig. 4**). In addition, the α -domains are also within cross-linking distance to the adjacent RbcL₂ unit in the RbcL₈–Raf1₄ complex (**Fig. 5h**). The resulting tentative model for the RbcL₈–Raf1₄ complex (**Fig. 5g,h**) is consistent with a role of Raf1 in stabilizing RbcL₂ and allowing its assembly into the RbcL₈ core complex. In the structure of RbcL₈ determined by cryo-EM, the ~60 C-terminal residues of RbcL

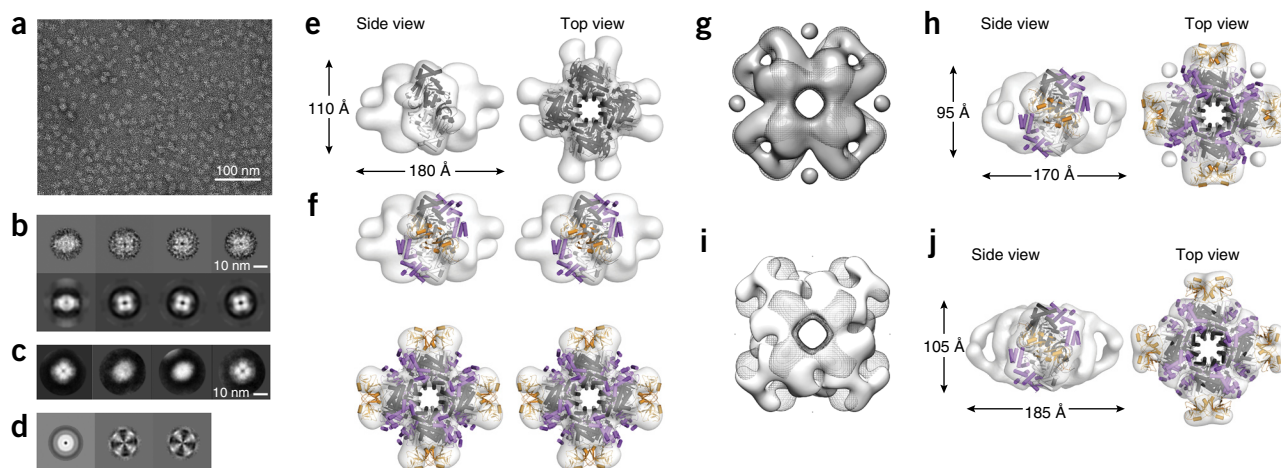


Figure 6 Negative-stain EM and 3D reconstructions of RbcL₈-Raf1₄ complex. **(a)** Micrograph of negatively stained complexes of *S. elongatus* RbcL₈ and Syn7942-Raf1 complex. **(b)** Class averages of the complexes derived from multivariate statistical analysis (MSA) in IMAGIC (upper row) and corresponding reprojections (bottom rows) of the initial 3D reconstruction. Each class average contains ~30 particles. **(c)** Class averages of the complexes from two-dimensional classification in RELION-1.3. **(d)** Eigenimages derived from MSA of top views in IMAGIC. **(e)** Rigid-body domain fitting of *S. elongatus* RbcL₈ into the final 3D reconstruction of RbcL₈-Raf1₄. Side and top views are shown. RbcL subunits are in gray and black. **(f)** Rigid-body domain fitting of *S. elongatus* RbcL₈ and the Syn7942-Raf1 α - and β -domains into the final 3D reconstruction of RbcL₈-Raf1₄. Side and top views are shown in stereo views. Gray and black, RbcL subunits; purple, Raf1 α ; orange, Raf1 β . **(g)** Negative-stain EM density of *S. elongatus* RbcL₈-Raf1₄ (dark mesh) overlaid on EM density of the same cross-linked complex (dark gray surface) in top view; the contour level is set to enclose 740 kDa. **(h)** Structural model of RbcL₈ and the Syn7942-Raf1 α - and β -domains docked into the cross-linked *S. elongatus* RbcL₈-Raf1₄ EM map shown as in **f**. **(i)** Negative-stain EM density of *S. elongatus* RbcL₈-Raf1₄ (dark mesh) overlaid on EM density of the *T. elongatus* RbcL₈-Raf1₄ complex (white surface) in top view. **(j)** Structural model of RbcL₈ and the Syn7942-Raf1 α - and β -domains docked into the *T. elongatus* RbcL₈-Raf1₄ EM map shown as in **f**.

are disordered but are ordered in the complex with the assembly chaperone RbcX^{14,22}. Thus, it seems plausible that the C-terminal residues of RbcL are also ordered in the complex with Raf1.

To obtain additional structural information, we next performed negative-stain EM and single-particle image analysis of the *S. elongatus* RbcL₈-Raf1₄ complex. Our reference-free analysis of two-dimensional class averages (7,602 particles) revealed four-fold symmetry in top views (**Fig. 6a–c**) and eigenimages from end views of the complex (**Fig. 6d**). Further analysis of a final data set of ~6,200 particles resulted in a 25-Å density map when dihedral four-fold symmetry was imposed (**Fig. 6e** and **Supplementary Fig. 7a**). When rendered at a threshold of 740-kDa mass, the particle dimensions were 110 × 180 Å. Compared to RbcL₈ in the holoenzyme crystal structure²⁸, the RbcL₈-Raf1₄ complex was ~10 Å taller and ~70 Å wider. However, it was ~35 Å taller than the cryo-EM structure of RbcL₈ alone, in which the C-terminal ~60 residues of RbcL are disordered¹⁴.

To determine the position of Raf1 in the complex, we first fitted the RbcL₈ core from the holoenzyme structure into the density with Chimera²⁹ (**Fig. 6e**). The Raf1 β dimer was docked into the protruding densities at the equatorial rim, and the α -domains were placed into additional densities at the top and bottom edges of RbcL₂ units (**Fig. 6f**). The cross-links mapped on this model with plausible distances (**Supplementary Data Set 1b** and **Supplementary Note**). Overall the EM- and CXMS-derived structural models are in reasonable agreement (**Supplementary Fig. 7b**) with the differences in domain orientations, thus possibly reflecting the dynamic nature of the interaction. We note that we were able to obtain a somewhat improved fit into the density when assuming that the ~60 C-terminal residues in RbcL are disordered (**Supplementary Fig. 7c**, **Supplementary Data Set 1b** and **Supplementary Note**). However, in this model the conserved surface of Raf1 α faced the solvent, and the distance

between the Raf1 α - and β -domains would be beyond the contour length of the linker residues.

To potentially stabilize the RbcL₈-Raf1₄ complex, we performed the EM analysis after DSS cross-linking (**Supplementary Fig. 7a,d**). The three-dimensional (3D) reconstruction obtained from 5,183 particles resembled the non-cross-linked complex, with a cross-correlation coefficient of 0.9834 (**Fig. 6g**). Although the cross-linked complex was slightly smaller (95 × 170 Å), and the Raf1 β dimer was stabilized (**Fig. 6h**), cross-linking did not reduce heterogeneity in the EM data set (**Supplementary Fig. 7d**), thus suggesting that structural dynamics is an intrinsic functional property of the RbcL₈-Raf1 interaction.

To further validate the EM structural model, we analyzed the RbcL-Raf1 complex of the thermophilic cyanobacterium *Thermosynechococcus elongatus* (**Supplementary Fig. 7a,e**). The purified complex, obtained upon coexpression of *T. elongatus* RbcL and Raf1 in *E. coli*, again contained four Raf1 dimers bound to RbcL₈ (**Supplementary Fig. 8**). The EM density of the *T. elongatus* RbcL₈-Raf1₄ complex was closely similar to that of *S. elongatus* (cross-correlation coefficient of 0.9742) (**Fig. 6i**), although its central pore was slightly wider (**Fig. 6j**).

Together the cross-linking data and the EM reconstructions support a model in which Raf1 brackets the RbcL antiparallel dimer, thus stabilizing it in a state competent for assembly to higher oligomers up to RbcL₈-Raf1₄.

DISCUSSION

Assembly of oligomeric protein complexes is widely considered to be a spontaneous process, and relatively little is known about the machineries that support the formation of specific multiprotein complexes. The biogenesis of hexadecameric Rubisco has emerged as a paradigm of assisted assembly^{30,31}. Here we analyzed the structure

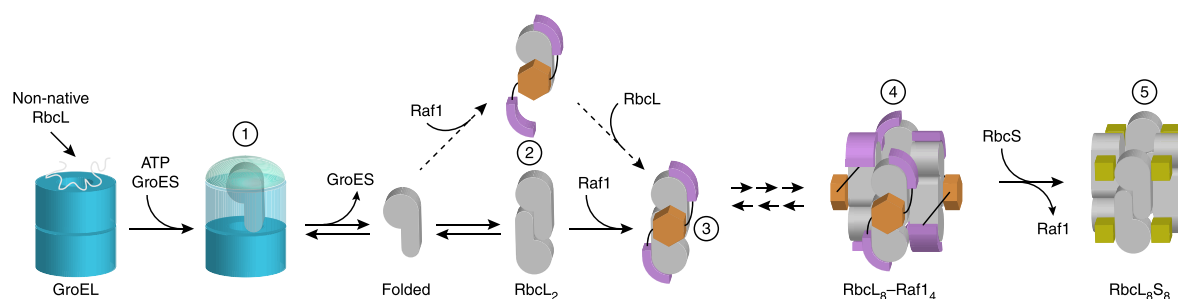


Figure 7 Model of Raf1-mediated Rubisco assembly. After folding by GroEL–ES (1), the released RbcL subunits either dimerize spontaneously or do so with assistance by Raf1 (2). The antiparallel RbcL dimer is stabilized by Raf1 (3) and is in dynamic equilibrium with higher oligomers up to RbcL₈–Raf1₄ complexes (4). The RbcL–Raf1 complex is dynamic, thus allowing RbcS binding to shift the equilibrium toward holoenzyme formation (5).

and mechanism of the Rubisco-assembly chaperone Raf1 (ref. 24). Our results demonstrate that the dimeric Raf1 functions downstream of RbcL-subunit folding by the GroEL–ES chaperonin system. The major intermediate populated during assembly is a complex in which Raf1 brackets the antiparallel RbcL dimer (Fig. 7), thus preventing rebinding of the structurally labile RbcL to GroEL. It is also possible that Raf1 brings RbcL subunits into proximity, thereby promoting dimer formation. The RbcL₂–Raf1 complex is competent for assembly into higher oligomeric states, with the RbcL₈–Raf1₄ complex as the endpoint (Fig. 7). Finally, binding of RbcS displaces Raf1 and completes assembly of the holoenzyme.

Like the structurally unrelated Rubisco-assembly chaperone RbcX^{16,32,33}, Raf1 is dimeric and engages in bivalent interactions with RbcL, a principle that probably relates to the antiparallel RbcL dimer being the building block of the RbcL₈ complex. As shown by X-ray crystallographic analysis, Raf1 consists of an N-terminal α -domain, a flexible linker segment and a C-terminal β -sheet domain that mediates dimerization. Both domains expose conserved interaction surfaces. From crystal structures, biochemical analysis, chemical cross-linking and negative-stain EM reconstruction, we propose a model of the RbcL₈–Raf1₄ assembly intermediate. In this model, the α -domains mediate the majority of functionally important contacts with RbcL by bracketing each RbcL dimer at the top and bottom, and the β -domain dimers are positioned coaxially in front of the RbcL₂ units. Because the α -domain alone is essentially inactive, dimerization is critical for Raf1 to achieve the necessary avidity for complex formation with RbcL assembly intermediates. At the same time, the interaction must remain dynamic in order to allow RbcS binding, because the Raf1 α -domain and RbcS have overlapping binding sites on RbcL.

Interestingly, the structurally distinct Raf1 and RbcX assembly proteins perform the same function, but they do so by using different interaction sites on RbcL. The boomerang-shaped RbcX dimer binds the C-terminal peptide tail of one RbcL subunit in a central cleft and contacts the N-terminal domain of the adjacent RbcL subunit via a peripheral region^{16,22}. In contrast, the Raf1–RbcL interaction surface appears to be more extensive. An interesting question therefore concerns whether Raf1 and RbcX act in parallel assembly pathways or functionally cooperate *in vivo* to achieve efficient assembly at a biologically relevant timescale. Taking into consideration that the binding sites for RbcX and RbcS have little if any overlap²², the previously described RbcL₈–RbcX₈ complex^{14,22} may be an additional assembly intermediate, especially when insufficient RbcS is present. A cooperation of Raf1 and RbcX, rather than mere functional redundancy, would be consistent with the strict co-occurrence of the two factors in photosynthetic organisms. Given that the RbcL subunits

of all form I Rubiscos are highly homologous, it is surprising that the Rubisco of higher plants has so far been refractory to reconstitution in *E. coli* or *in vitro*. Perhaps additional factors are required for the assembly of the plant enzyme^{24,34,35}.

The chaperone requirement for the folding and assembly of hexadecameric (form I) Rubisco is remarkably complex, thus raising the question of why such an important and abundant enzyme has failed to evolve a more robust biogenesis pathway. Form I Rubisco evolved from the simpler dimeric form II enzyme about 2.5 billion years ago to adapt to the increasing levels of oxygen³⁶. Both the form I and form II RbcL subunits require the GroEL–ES chaperonin for folding, like many other TIM-barrel proteins^{37,38}. However, only the RbcL of form I Rubisco remains structurally labile after folding and tends to rebinding to chaperonin rather than to spontaneously assemble. The dependence on assembly chaperones, in turn, is likely to have slowed the evolution of form I Rubisco. Recent attempts to improve the enzymatic properties of the enzyme by directed evolution have shown that although the chaperonin tends to increase the number of permissible mutations, the assembly chaperone RbcX retards evolvability¹¹. It remains to be seen whether Raf1 confers similar evolutionary constraints.

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Accession codes. Coordinates for the structural domains of the protein Raf1.2 from *A. thaliana* have been deposited in the Protein Data Bank under accession codes **4WT3**, **4WT4** and **4WT5**. Negative-stain EM maps of the *S. elongatus* and *T. elongatus* RbcL₈–Raf1₄ complexes have been deposited in the Electron Microscopy Data Bank under accession codes **EMD-3051** (*T. elongatus* RbcL₈–Raf1₄), **EMD-3052** (cross-linked *S. elongatus* RbcL₈–Raf1₄) and **EMD-3053** (*S. elongatus* RbcL₈–Raf1₄).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

A. thaliana cDNA was a kind gift from B. Bölter (Ludwig-Maximilians-Universität München), and *S. elongatus* PCC7942 DNA was a kind gift from M. Hagemann (Universität Rostock). We thank A. Jungclaus for assistance with protein purification and R. Körner for performing the MS on the cross-linked samples. We also thank A. Sinz (Martin Luther University Halle-Wittenberg) for making available StavroX software for the cross-linking MS analysis. Assistance by K. Valer and J. Basquin at the Max Planck Institute of Biochemistry (MPIB) crystallization facility, as well as the staff at Swiss Light Source beamlines X10SA-PX-II and X06DA-PXIII, is gratefully acknowledged. SAXS data were collected at European

Synchrotron Radiation Facility beamline BM29 with the assistance of A. Round. Cryo-EM analysis was performed in the Department of W. Baumeister at MPIB. The authors thank A. Kahraman (Universität Zürich) for helpful discussions regarding the interpretation of cross-linking data and D. Balchin and L. Popilka for critically reading the manuscript. P.W. is supported by an Emmy Noether grant of the German Research council (WE 4628/1).

AUTHOR CONTRIBUTIONS

T.H. performed the biochemical and functional analysis of Rbf1 and obtained the Rbf1-domain crystals; A.B. solved the crystal structures; J.Y.B. performed the native MS and analyzed the cross-linking MS data; M.H.-H. did the SEC-MALS measurements; and G.M. and P.W. performed the cryo-EM analysis and reconstruction. M.H.-H., F.U.H. and A.B. supervised the experimental design and data interpretation, and wrote the manuscript with contributions from all authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Andersson, I. & Backlund, A. Structure and function of Rubisco. *Plant Physiol. Biochem.* **46**, 275–291 (2008).
- Hartman, F.C. & Harpel, M.R. Structure, function, regulation, and assembly of D-ribulose-1,5-bisphosphate carboxylase/oxygenase. *Annu. Rev. Biochem.* **63**, 197–234 (1994).
- Ellis, R.J. Biochemistry: tackling unintelligent design. *Nature* **463**, 164–165 (2010).
- Sharkey, T.D. Estimating the rate of photorespiration. *Physiol. Plant.* **73**, 147–152 (1988).
- Portis, A.R. & Parry, M.A.J. Discoveries in Rubisco (ribulose 1,5-bisphosphate carboxylase/oxygenase): a historical perspective. *Photosynth. Res.* **94**, 121–143 (2007).
- Peterhansel, C., Niessen, M. & Kebeish, R.M. Metabolic engineering towards the enhancement of photosynthesis. *Photochem. Photobiol.* **84**, 1317–1323 (2008).
- Maurino, V.G. & Peterhansel, C. Photorespiration: current status and approaches for metabolic engineering. *Curr. Opin. Plant Biol.* **13**, 249–256 (2010).
- Whitney, S.M., Houtz, R.L. & Alonso, H. Advancing our understanding and capacity to engineer nature's CO₂-sequestering enzyme, Rubisco. *Plant Physiol.* **155**, 27–35 (2011).
- Parry, M.A.J. *et al.* Rubisco activity and regulation as targets for crop improvement. *J. Exp. Bot.* **64**, 717–730 (2013).
- Lin, M.T., Occhialini, A., Andralojc, P.J., Parry, M.A. & Hanson, M.R. A faster Rubisco with potential to increase photosynthesis in crops. *Nature* **513**, 547–550 (2014).
- Durão, P. *et al.* Opposing effects of folding and assembly chaperones on evolvability of Rubisco. *Nat. Chem. Biol.* **11**, 148–155 (2015).
- Cleland, W.W., Andrews, T.J., Gutteridge, S., Hartman, F.C. & Lorimer, G.H. Mechanism of Rubisco: the carbamate as general base. *Chem. Rev.* **98**, 549–562 (1998).
- Andersson, I. Catalysis and regulation in Rubisco. *J. Exp. Bot.* **59**, 1555–1568 (2008).
- Liu, C. *et al.* Coupled chaperone action in folding and assembly of hexadecameric Rubisco. *Nature* **463**, 197–202 (2010).
- Vitlin Gruber, A., Nisemblat, S., Azem, A. & Weiss, C. The complexity of chloroplast chaperonins. *Trends Plant Sci.* **18**, 688–694 (2013).
- Saschenbrecker, S. *et al.* Structure and function of RbcX, an assembly chaperone for hexadecameric Rubisco. *Cell* **129**, 1189–1200 (2007).
- van der Vies, S.M., Bradley, D. & Gatenby, A.A. Assembly of cyanobacterial and higher plant ribulose bisphosphate carboxylase subunits into functional homologous and heterologous enzyme molecules in *Escherichia coli*. *EMBO J.* **5**, 2439–2444 (1986).
- Andrews, T.J. Catalysis by cyanobacterial ribulose-bisphosphate carboxylase large subunits in the complete absence of small subunits. *J. Biol. Chem.* **263**, 12213–12219 (1988).
- Goloubinoff, P., Gatenby, A.A. & Lorimer, G.H. GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in *Escherichia coli*. *Nature* **337**, 44–47 (1989).
- Li, L.-A. & Tabita, F.R. Maximum activity of recombinant ribulose 1,5-bisphosphate carboxylase/oxygenase of *Anabaena* sp. strain CA requires the product of the *rbcx* gene. *J. Bacteriol.* **179**, 3793–3796 (1997).
- Onizuka, T. *et al.* The *rbcX* gene product promotes the production and assembly of ribulose-1,5-bisphosphate carboxylase/oxygenase of *Synechococcus* sp. PCC7002 in *Escherichia coli*. *Plant Cell Physiol.* **45**, 1390–1395 (2004).
- Bracher, A., Starling-Windhof, A., Hartl, F.U. & Hayer-Hartl, M. Crystal structure of a chaperone-bound assembly intermediate of form I Rubisco. *Nat. Struct. Mol. Biol.* **18**, 875–880 (2011).
- Emlyn-Jones, D., Woodger, F.J., Price, G.D. & Whitney, S.M. RbcX can function as a Rubisco chaperonin, but is non-essential in *Synechococcus* PCC7942. *Plant Cell Physiol.* **47**, 1630–1640 (2006).
- Feiz, L. *et al.* Ribulose-1,5-bis-phosphate carboxylase/oxygenase accumulation factor1 is required for holoenzyme assembly in maize. *Plant Cell* **24**, 3435–3446 (2012).
- Kolesinski, P., Belusiak, I., Czarnocki-Cieciura, M. & Szczepaniak, A. Rubisco accumulation factor 1 from *Thermosynechococcus elongatus* participates in the final stages of ribulose-1,5-bisphosphate carboxylase/oxygenase assembly in *Escherichia coli* cells and *in vitro*. *FEBS J.* **281**, 3920–3932 (2014).
- Whitney, S.M., Birch, R., Kelso, C., Beck, J.L. & Kapralov, M.V. Improving recombinant Rubisco biogenesis, plant photosynthesis and growth by coexpressing its ancillary RAF1 chaperone. *Proc. Natl. Acad. Sci. USA* **112**, 3564–3569 (2015).
- Leitner, A. *et al.* The molecular architecture of the eukaryotic chaperonin TRiC/CCT. *Structure* **20**, 814–825 (2012).
- Newman, J., Bränden, C.I. & Jones, T.A. Structure determination and refinement of ribulose 1,5-bisphosphate carboxylase/oxygenase from *Synechococcus* PCC6301. *Acta Crystallogr. D Biol. Crystallogr.* **49**, 548–560 (1993).
- Pettersen, E.F. *et al.* UCSF chimera: a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
- Chari, A. & Fischer, U. Cellular strategies for the assembly of molecular machines. *Trends Biochem. Sci.* **35**, 676–683 (2010).
- Ellis, R.J. Assembly chaperones: a perspective. *Phil. Trans. R. Soc. Lond. B* **368**, 20110398 (2013).
- Tanaka, S., Sawaya, M.R., Kerfeld, C.A. & Yeates, T.O. Structure of the RuBisCO chaperone RbcX from *Synechocystis* sp. PCC6803. *Acta Crystallogr. D Biol. Crystallogr.* **63**, 1109–1112 (2007).
- Kolesinski, P. *et al.* Insights into eukaryotic Rubisco assembly: crystal structures of RbcX chaperones from *Arabidopsis thaliana*. *Biochim. Biophys. Acta* **1830**, 2899–2906 (2013).
- Wheatley, N.M., Sundberg, C.D., Gidaniyan, S.D., Cascio, D. & Yeates, T.O. Structure and identification of a pterin dehydratase-like protein as a ribulose-bisphosphate carboxylase/oxygenase (RuBisCO) assembly factor in the alpha-carboxysome. *J. Biol. Chem.* **289**, 7973–7981 (2014).
- Feiz, L. *et al.* RAF2: a novel Rubisco biogenesis factor in maize. *Plant J.* **80**, 862–869 (2014).
- Hohmann-Marriott, M.F. & Blankenship, R.E. Evolution of photosynthesis. *Annu. Rev. Plant Biol.* **62**, 515–548 (2011).
- Kerner, M.J. *et al.* Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell* **122**, 209–220 (2005).
- Georgescauld, F. *et al.* GroEL/ES chaperonin modulates the mechanism and accelerates the rate of TIM-barrel domain folding. *Cell* **157**, 922–934 (2014).

ONLINE METHODS

Cloning and plasmids. Open reading frames for Syn7942-Raf1, Syn7002-Raf1, AtRaf1.1 and AtRaf1.2 were amplified by PCR from genomic DNA of *S. elongatus* PCC7942, *Synechococcus* sp. PCC7002 (ATCC no. 27264) and *A. thaliana* cDNA, respectively, and cloned between the SacII and SacI restriction sites of the pHue plasmid³⁹, thus resulting in the following constructs: pHueSyn7942raf1, pHueSyn7002raf1, pHueAtraf1.1 and pHueAtraf1.2. The chloroplast transit peptides of AtRaf1.1 and AtRaf1.2 were predicted with TargetP (<http://www.cbs.dtu.dk/services/TargetP/>) or adopted from the plant proteome database (<http://ppdb.tc.cornell.edu/>). The bicistronic plasmid pHueAtraf1.1raf1.2 was created by amplifying Atraf1.2 from pHueAtraf1.2 and inserting it into pHueAtraf1.1 with SacI and NotI restriction sites. The respective primer sequences are listed in **Supplementary Table 1**. SacII and SacI sites in the protein-coding regions were removed with whole-plasmid site-directed mutagenesis. The sequences for the N-terminal (Raf1 α) and C-terminal (Raf1 β) Raf1 domains were analogously cloned into pHue. *T. elongatus* BP-1 genes encoding RbcL and Raf1, and were synthesized (Life Technologies) and cloned between the NcoI and NotI and NdeI and XhoI restriction sites of the bicistronic pCDF-Duet-1 plasmid, respectively (Novagen). Raf1 was synthesized to contain a TEV protease-cleavable N-terminal His-tag (MGSSHHHHHHENLYFQG). For FLAG-tagged constructs, the sequence encoding the FLAG tag (MDYKDDDDKAG) was inserted upstream of the respective start codon (as described above). Point mutants were produced by PCR-based mutagenesis. All plasmid inserts were verified by DNA sequencing.

Protein expression and purification. All purification steps were performed at 4 °C, and protein concentrations were determined spectrophotometrically at 280 nm. Raf1 proteins were expressed as N-terminal His₆-ubiquitin fusion proteins in *E. coli* BL21(DE3) cells containing the given pHue expression plasmid. Cells were grown to an OD₆₀₀ of 0.5 at 37 °C in Luria-Bertani medium; this was followed by induction for 16 h with 0.5 mM isopropyl β -D-thiogalactoside at 23 °C. Cells were lysed in 50 mM Tris-HCl, pH 8.0, 20 mM NaCl, 1 mM EDTA, 0.5 mg mL⁻¹ lysozyme and 5 mM phenylmethylsulfonyl fluoride for 30 min on ice, and this was followed by ultrasonication (Misonix Sonicator 3000). The supernatant obtained after high-speed centrifugation (48,000g, 40 min, 4 °C) was applied to an Ni-IMAC column (GE Biotech) to capture the His₆-ubiquitin-Raf1 fusion protein. This was followed by overnight cleavage of the His₆-ubiquitin moiety at 23 °C with the deubiquitinating enzyme Usp2 (ref. 40). The protein-containing fraction was dialyzed against buffer A (20 mM Tris-HCl, pH 8.0, and 50 mM NaCl) and applied to a MonoQ (GE Biotech) column equilibrated with buffer A. Proteins were eluted with a linear salt gradient to 1 M NaCl. Fractions containing Raf1 were combined and concentrated, 5% glycerol was added, and fractions were flash frozen in liquid nitrogen and stored at -80 °C. Raf1 for X-ray crystallographic studies was purified further by Superdex200 (GE Biotech) size-exclusion chromatography in buffer A.

T. elongatus RbcL-Raf1 complex was expressed and purified essentially as described for Raf1 proteins. After purification with an Ni-IMAC column, the complex was dialyzed against buffer A. This was followed by purification with an ion-exchange column (MonoQ) and a final size-exclusion-chromatography step in buffer A (Superdex200). Fractions containing the HMW *T. elongatus* RbcL-Raf1 complex were combined and concentrated, 5% glycerol was added, and fractions were flash frozen in liquid nitrogen and stored at -80 °C.

S. elongatus 6301 (Syn6301) RbcL₈S₈, RbcL₈ and RbcS, as well as unassembled RbcL, were purified as previously described^{14,16}. The Rubisco proteins of Syn6301 and Syn7942 are 100% identical in sequence (UniProt P00880, Q31NB3, P04716 and Q31NB2). Rabbit antibodies against *S. elongatus* RbcL (ref. 14), Syn7942-Raf1 and AtRaf1.2 were produced with standard procedures (**Supplementary Data Set 2**). The *E. coli* chaperonins GroEL and GroES were purified as described previously⁴¹.

Rubisco reconstitution. GroEL-ES-mediated RbcL folding was performed as in ref. 14, with modifications. Denatured *S. elongatus* RbcL was diluted 200-fold from 6 M GuHCl (final concentration 0.5 μ M) into ice-cold buffer B (20 mM MOPS-KOH, pH 7.5, 100 mM KCl, and 5 mM MgOAc₂) containing 5 mM DTT, 1 mg mL⁻¹ BSA and 1 μ M GroEL, and was incubated on ice and then centrifuged to remove any aggregated protein. GroES (2 μ M), *S. elongatus* RbcS (5 μ M) and Raf1 were added to the supernatant containing GroEL-bound RbcL when indicated in figure legends. Folding and assembly was initiated by addition of 4 mM ATP and

incubation at 25 °C. Reactions were stopped at the indicated times by addition of apyrase (Sigma) to a final concentration of 0.25 U μ L⁻¹ to inhibit GroEL-ES activity. (Apyrase hydrolyzes ATP and ADP to AMP.) Formation of assembled complexes was analyzed by continuous Tris-acetate native-PAGE gradient gels (5–15% acrylamide) and immunoblotting with anti-RbcL and anti-Raf1 antibodies.

For measurement of Rubisco enzymatic activity, aliquots of the folding assay were supplemented with 5 μ M *S. elongatus* RbcS when it had not previously been present in the reaction, and RbcL₈S₈ assembly was allowed to proceed for 15 min at 25 °C. Rubisco carboxylation activity was determined after incubation at 25 °C for 10 min in 50 mM Tris-HCl, pH 8.0, 10 mM MgCl₂, and 30 mM NaH¹⁴CO₃ (25 Bq/nmol), and the amount of fixed carbon was quantified as described previously⁴¹. Activities are expressed as a percentage of purified *S. elongatus* RbcL₈ standard supplemented with *S. elongatus* RbcS (control).

Interaction of Raf1 with preformed RbcL₈ complexes. Purified RbcL₈ and RbcL₈S₈ complexes were incubated with Raf1 for 15 min at 25 °C in buffer B containing 5 mM DTT. Formation of assembled complexes was analyzed by continuous Tris-acetate native PAGE (5–15% acrylamide) and immunoblotting with anti-RbcL and anti-Raf1 antibodies and by SEC-MALS and native MS.

Size-exclusion chromatography coupled to multi-angle static light scattering (SEC-MALS). Protein samples (~30 to 60 μ g) were analyzed with static and dynamic light scattering by autoinjection of the sample onto a SEC column (5 μ m, 7.8 \times 300 mm column, Wyatt Technology WTC-030N5) at a flow rate of 0.2 mL min⁻¹ in buffer C (25 mM Tris-HCl, pH 8.0, and 50 mM NaCl) at 25 °C or at a flow rate of 0.15 mL min⁻¹ in buffer B for analysis of the RbcL₈-Raf1 complexes in solution. The column is inline with the following detectors: a variable UV-absorbance detector set at 280 nm (Agilent 1100 series), a DAWN EOS MALS detector (Wyatt Technology, 690-nm laser) and an Optilab rEX refractive-index detector (Wyatt Technology, 690-nm laser)⁴². Masses were calculated with ASTRA (Wyatt Technology) with the dn/dc value set to 0.185 mL g⁻¹. Bovine serum albumin (Thermo) was used as the calibration standard.

Native mass spectrometry. The purified *S. elongatus* RbcL₈ (1.25 μ M oligomer), Syn7942-Raf1 (5 μ M dimer) and the binding reaction (15 min at 25 °C) of Syn7942-Raf1 (5 μ M dimer) to RbcL₈ (1.25 μ M oligomer) in buffer D (20 mM MOPS-KOH, pH 7.5, 50 mM KCl, and 5 mM MgOAc₂) containing 5 mM DTT were buffer-exchanged into 100 mM ammonium acetate, pH 7.5–8.0 (Fluka, Sigma) with Micro Bio-Spin 6 chromatography columns (Bio-Rad). The purified *T. elongatus* RbcL-Raf1 complex (~8 μ M) was buffer-exchanged as above. Native-MS analyses were performed in positive-ion mode on an electrospray ionization quadrupole time-of-flight (ESI-TOF) hybrid mass spectrometer (Synapt G2-Si, Waters) equipped with a Z-spray nano-ESI source. Gold-plated 10- μ m nano-ESI pipettes (Mascom) were used as capillaries. Optimized capillary and sample cone voltages were 1–1.6 kV and 100–150 V, respectively. Spectra were calibrated with 30 mg mL⁻¹ cesium iodide dissolved in 1:1 acetonitrile/water.

Electron microscopy (EM) and 3D reconstruction. RbcL₈-Raf1₄ complex (~50 μ g mL⁻¹) in buffer B containing 5 mM DTT was applied to freshly plasma-cleaned, carbon-coated grids (Quantifoil) and was negative stained with 2% (w/v) uranyl acetate. Images were digitally recorded on a Philips CM200 FEG electron microscope equipped with a TemCam F415MP 4k at a nominal magnification of 50,000 \times , with a pixel size of 2.16 Å/pixel at specimen level and defocus ranging from 300 to 1,500 nm. The microscope was operated under low-dose conditions at 160 kV. The defocus and astigmatism of the images were determined with CTFFIND4 (ref. 43) in RELION-1.3 (ref. 44). A total of 62, 49 and 90 micrographs were selected for image processing of the *S. elongatus* RbcL₈-Raf1₄, the cross-linked *S. elongatus* RbcL₈-Raf1₄ and the *T. elongatus* RbcL₈-Raf1₄ complexes, respectively. Manual particle selection with RELION-1.3 yielded a final data set of 5,183 particles for the cross-linked *S. elongatus* RbcL₈-Raf1₄ complex and 5,471 particles for the *T. elongatus* RbcL₈-Raf1₄ complex. For the native *S. elongatus* RbcL₈-Raf1₄ complex, a total of 1,057 particles from 13 micrographs were manually selected with RELION-1.3 to generate references for automatic particle picking. After automatic particle selection in RELION-1.3 and manual cleaning of the data set on the basis of z-score characteristics, a stack of 7,602 particles was obtained. The extracted particle data set was subjected to 2D classification in RELION-1.3, and classes showing erroneously picked features

and aggregates were discarded, thus resulting in 6,191 particles in the final data set. Initial 3D reconstruction of the *S. elongatus* RbcL₈-Raf1₄ complex and the *T. elongatus* RbcL₈-Raf1₄ complex were generated with IMAGIC-5 (ref. 45). Particle images were band-pass-filtered between 250 and 10 Å, normalized and centered by iterative alignment to their rotationally averaged sum. Initial class averages containing ~30 particles were obtained by two rounds of classification on the basis of multivariate statistical analysis (MSA) and subsequent multireference alignment with homogenous classes as new references. For symmetry analysis, top views of the complexes were extracted, randomly rotated and subjected to MSA. A density map with imposed dihedral four-fold symmetry was created by angular reconstitution. 3D refinement of the initial model was carried out in RELION-1.3 with application of four-fold symmetry and no mask. The resolution of the final refinement by gold-standard Fourier shell correlation (FSC) at cutoffs of 0.143 and 0.5 is shown in **Supplementary Figure 7a**. The PDB structures of RbcL₈ (from the Rubisco crystal structure²⁸) and Raf1 domains were fitted manually as rigid bodies with UCSF Chimera²⁹.

Cross-linking coupled to mass spectrometry (CXMS). Purified RbcL₈ (1.25 μM oligomer) was incubated with Raf1 (10 μM) in buffer D for 15 min at 25 °C. This was followed by addition of 2 mM of a 1:1 isotopic mixture of H₁₂ and D₁₂ disuccinimidylsuberate, H₁₂/D₁₂-DSS (Creative Molecules) for 30 min (**Supplementary Fig. 6**). Reactions were quenched by addition of 150 mM NH₄HCO₃. (CXMS) with the bifunctional lysine-specific cross-linker disuccinimidylsuberate (DSS). Cross-linking during reconstitution was performed with N-terminally FLAG-tagged Raf1; this was followed by isolation of cross-linked Raf1 by anti-FLAG immunoprecipitation (anti-FLAG beads, Sigma). Cross-linked complexes were separated on 4–12% Bis-Tris SDS-PAGE (Life Technologies) and visualized by Coomassie staining. Excision of gel bands, reduction, alkylation, in-gel digestion and desalting were performed as previously described^{46,47}. Desalted peptides were dissolved in 5% formic acid and analyzed on an Easy-nLC 1000 UPLC system (Thermo) connected to a Q-Exactive Orbitrap mass spectrometer (Thermo). Peptides were separated on nanospray columns (ID 75 μm, 20 cm long with 8-μm tip opening, New Objective) containing 1.9-μm C18 beads (Reprosil-Pur C18-AQ, Dr. Maisch) on a 60-min gradient with buffers (0.2% formic acid in water and 0.2% formic acid in acetonitrile). Sample loading to the column was performed by the Thermo Easy-nLC 1000 autosampler without a trap column, at a flow rate of 0.5 μl per min. The UPLC flow rate during sample analysis was set to 0.2 μl/min. MS² analysis was performed with standard data-dependent mode settings, with alternation between one high-resolution (resolution 70,000) MS1 scan (*m/z* of 400–1,750) and ten MS2 scans (resolution 17,500) of the ten most intense ions with charge states of three or more, as run on Thermo Xcalibur software.

For data analysis, Thermo Xcalibur .raw files were converted to .mgf (Mascot generic file) format, with Proteome Discoverer 1.4 (Thermo). To identify cross-linked peptides, the .mgf files were analyzed on StavroX 3.1.19 (ref. 48) with the following parameters: MS1 tolerance of 3 p.p.m.; MS2 tolerance of 0.8 Da; missed cleavages for lysine and arginine of 3 and 1, respectively; signal-to-noise ratio ≥2; and fixed and variable modifications of cysteine carbamidomethylation and methionine oxidation, respectively. The potential cross-linking sites for DSS are lysine residues and the free amino group at the N terminus. All the identified linked peptides were validated manually for b- and y-ion assignment and were included in the final list only if precursors of the respective peptides contained doublets with mass difference of 12.0753 Da (mass difference between H₁₂-DSS and D₁₂-DSS). The cross-linked peptides were identified from two independent experiments for each Raf1 protein.

Small-angle X-ray scattering (SAXS). SAXS measurements were performed at beamline BM29 of the European Synchrotron Radiation Source (ESRF), Grenoble, France. Protein samples at three different concentrations in 50 mM Tris-HCl, pH 8.0, 50 mM NaCl and 5% (v/v) glycerol were exposed for 1 s. Scattering data from ten repeats were averaged. Buffer background was subtracted. The protein scattering data were processed with Primus^{49,50}. Radii of gyration were determined with the Guinier approximation. Scattering curves were fitted with GNOM⁵¹. An ensemble model of the AtRaf1.2 structure was generated with EOM 2.0 (refs. 52,53).

Crystallization. Initial crystals of AtRaf1.2α were obtained by the sitting-drop vapor-diffusion method at 18 °C, with mixture of equal volumes of AtRaf1.2(62–449)

(18 mg mL⁻¹ in buffer C) with a precipitant containing 0.1 M MES-NaOH, pH 6.0, and 20% (w/v) PEG 6000 (Qiagen pHClear Suite condition D3). Crystals of AtRaf1.2α(62–274) were obtained at 20 °C with 0.1 M MES-NaOH, pH 6.0, and 26% (w/v) PEG-6000 as a precipitant. For cryoprotection, crystals were transferred stepwise into 0.1 M MES-NaOH, pH 6.0, 26% (w/v) PEG 6000 and 15% (v/v) glycerol before being flash frozen in liquid nitrogen.

Orthorhombic crystals of AtRaf1.2β were obtained by the sitting-drop vapor-diffusion method at 4 °C, with mixture of equal volumes of AtRaf1.2(62–449) (17 mg mL⁻¹ in buffer C) with a precipitant containing 20% PEG 6000, 50 mM K₂HPO₄ and 200 mM KH₂PO₄. Monoclinic crystals of AtRaf1.2β(281–449) were obtained with 10% PEG3350 as a precipitant.

Structure determination. Diffraction data were collected at beamlines X10SA and X06DA of the Swiss Light Source (SLS) in Villigen, Switzerland. The data were processed with XDS⁵⁴ and transferred into the CCP4 format with Pointless⁵⁵, Scala⁵⁶ and Truncate⁵⁷. The structures of AtRaf1.2α and monoclinic AtRaf1.2β were solved by SIRAS with crystals soaked with 1 mM K₂[PtCl₄] and sodium *p*-(hydroxymercuro)benzoate (HMBA), respectively. Two platinum and six mercury sites were found with SHELXD⁵⁸, as implemented in HKL2MAP⁵⁹, for AtRaf1.2α and AtRaf1.2β, respectively. These solutions were further refined with Sharp⁶⁰. Density modification was subsequently carried out with Resolve⁶¹. A preliminary model for AtRaf1.2α was manually built in the resulting map with Coot⁶². For final model building and refinement, a nearly isomorphous native data of AtRaf1.2α(62–274) was used. Buccaneer⁶³ built a partial model of AtRaf1.2β, which was completed manually with Coot. Orthorhombic AtRaf1.2β was solved by molecular replacement with Molrep⁶⁴. The models were improved by iterative cycles of refinement with Refmac⁶⁵, as implemented in the CCP4 interface⁶⁶, and this was followed by manual model building. The final refinement of AtRaf1.2α was carried out with phenix.refine⁶⁷. Nonglycine residues facing solvent channels without discernible side chain density were modeled as alanines.

Structure analysis. Coordinates were aligned with Lsqman⁶⁸. The sequence alignment was prepared with ClustalW^{69,70} and ESPript⁷¹. Secondary-structure prediction was performed with Jpred-3 (ref. 72). Figures were generated with PyMOL (<http://www.pymol.org/>).

39. Catanzariti, A.M., Soboleva, T.A., Jans, D.A., Board, P.G. & Baker, R.T. An efficient system for high-level expression and easy purification of authentic recombinant proteins. *Protein Sci.* **13**, 1331–1339 (2004).
40. Baker, R.T. *et al.* Using deubiquitylating enzymes as research tools. *Methods Enzymol.* **398**, 540–554 (2005).
41. Brinker, A. *et al.* Dual function of protein confinement in chaperonin-assisted protein folding. *Cell* **107**, 223–233 (2001).
42. Wyatt, P.J. Light scattering and the absolute characterization of macromolecules. *Anal. Chim. Acta* **272**, 1–40 (1993).
43. Mindell, J.A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
44. Scheres, S.H. Semi-automated selection of cryo-EM particles in RELION-1.3. *J. Struct. Biol.* **189**, 114–122 (2015).
45. van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116**, 17–24 (1996).
46. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850–858 (1996).
47. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).
48. Götz, M. *et al.* Automated assignment of MS/MS cleavable crosslinks in protein 3D-structure analysis. *J. Am. Soc. Mass Spectrom.* **26**, 83–97 (2015).
49. Vachette, P., Koch, M.H. & Svergun, D.I. Looking behind the beamstop: X-ray solution scattering studies of structure and conformational changes of biological macromolecules. *Methods Enzymol.* **374**, 584–615 (2003).
50. Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J. & Svergun, D.I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 (2003).
51. Svergun, D.I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 (1992).
52. Bernadó, P., Mylonas, E., Petoukhov, M.V., Blackledge, M. & Svergun, D.I. Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* **129**, 5656–5664 (2007).
53. Petoukhov, M.V. *et al.* New developments in the program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350 (2012).
54. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
55. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82 (2006).

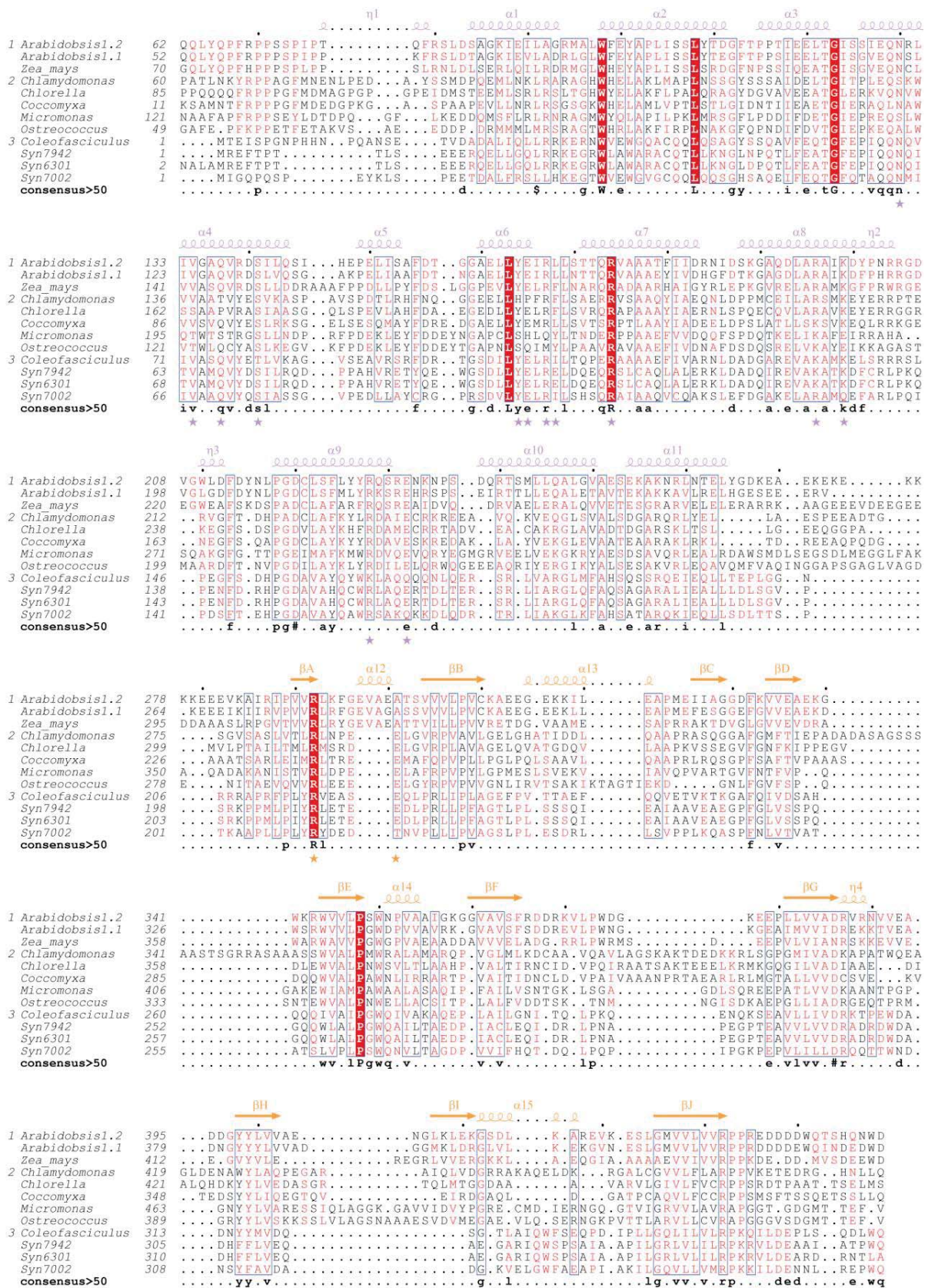


56. Evans, P.R. Scala. *Joint CCP4 ESF-EACBM Newsl. Prot. Crystallogr.* **33**, 22–24 (1997).
57. French, G. & Wilson, K. On the treatment of negative intensity observations. *Acta Crystallogr. A* **34**, 517–525 (1978).
58. Sheldrick, G.M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 479–485 (2010).
59. Pape, T. & Schneider, T.R. HKL2MAP: a graphical user interface for phasing with SHELX programs. *J. Appl. Crystallogr.* **37**, 843–844 (2004).
60. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230 (2007).
61. Terwilliger, T.C. Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 965–972 (2000).
62. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
63. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011 (2006).
64. Vagin, A.A. & Isupov, M.N. Spherically averaged phased translation function and its application to the search for molecules and fragments in electron-density maps. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 1451–1456 (2001).
65. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).
66. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763 (1994).
67. Afonine, P.V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).
68. Kleywegt, G.T. & Jones, T.A. A super position. *CCP4 ESF-EACBM Newsl. Prot. Crystallogr.* **31**, 9–14 (1994).
69. Thompson, J.D., Higgins, D.G. & Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
70. Larkin, M.A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
71. Gouet, P., Courcelle, E., Stuart, D.I. & Metz, F. ESPript: multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308 (1999).
72. Cole, C., Barber, J.D. & Barton, G.J. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.* **36**, W197–W201 (2008).

SUPPLEMENTARY INFORMATION

**Structure and mechanism of the Rubisco
assembly chaperone Raf1**

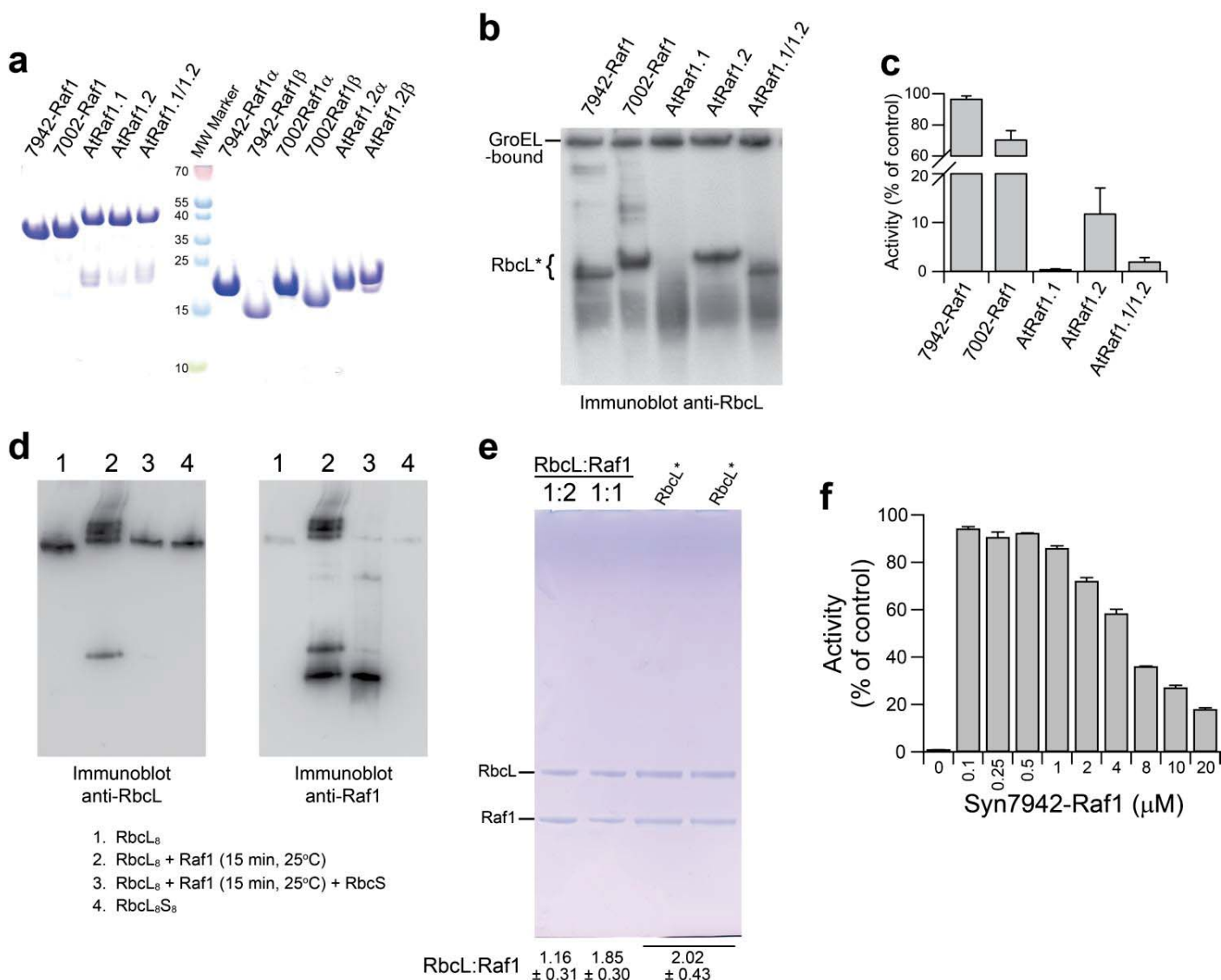
Thomas Hauser, Javaid Y. Bhat, Goran Miličić, Petra Wendler, F. Ulrich Hartl, Andreas Bracher
and Manajit Hayer-Hartl



Supplementary Figure 1

Alignment of Raf1 sequences.

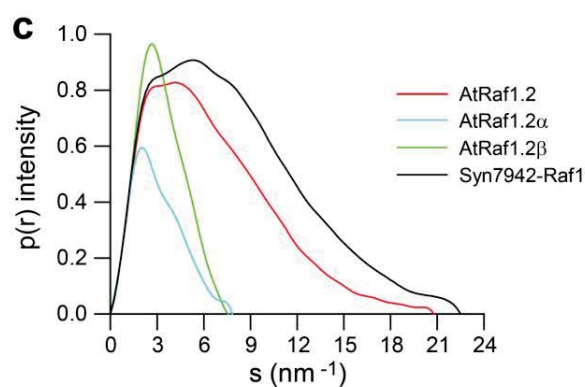
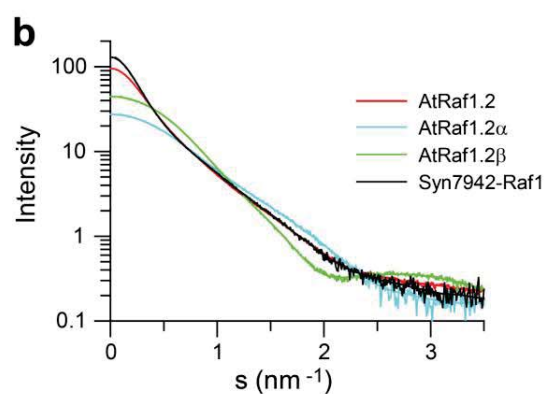
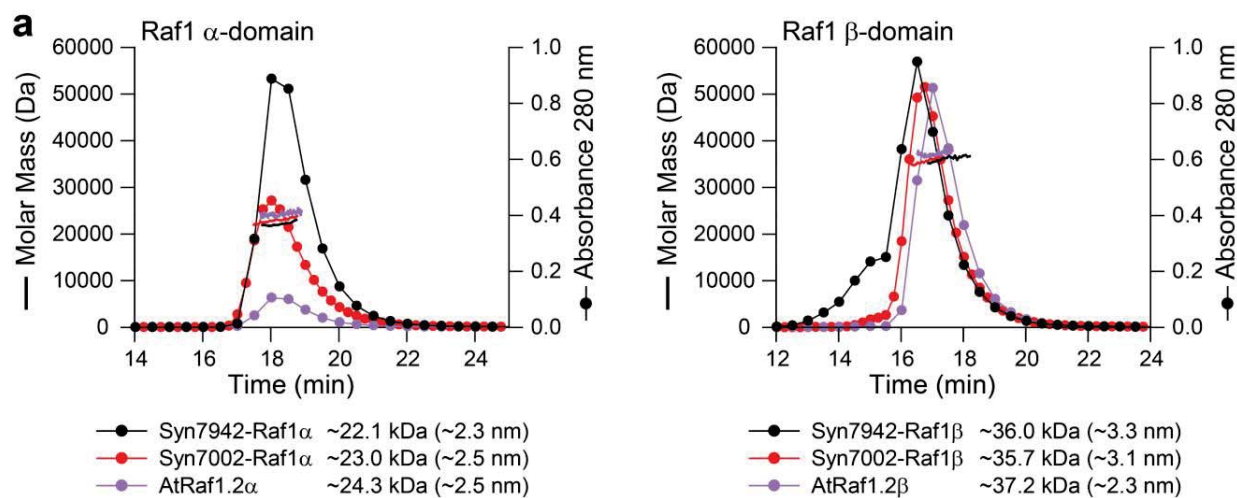
Amino acid sequences of a representative set of Raf1 homologs were aligned using the EBI Clustal-Ω server. Secondary structure elements for Raf1.2 from *Arabidopsis thaliana* are indicated above the sequences. The Raf1 domain structure is indicated by purple and orange coloring of secondary structure elements in the Raf1 α and Raf1 β domains, respectively. The sequences from plants (1), green algae (2) and cyanobacteria (3) are grouped separately. Similar residues are shown in red and identical residues in white on a red background. Blue frames indicate homologous regions. The consensus sequence is shown at the bottom. The chloroplast signal sequences are not shown. Asterisks below the sequence indicate mutations in Syn7942-Raf1 analyzed in this study (Fig. 4). The Uniprot accession codes for the sequences are: Q9SR19, *Arabidopsis thaliana* Raf1.2; Q9LKR8, *Arabidopsis thaliana* Raf1.1; B4FR29, *Zea mays*; I0YJW5, *Coccomyxa subellipsoidea* C-169; E1ZGR5, *Chlorella variabilis*; Cre06.g308450.t1.2, *Chlamydomonas reinhardtii*; Q00S02, *Ostreococcus tauri*; C1FI81, *Micromonas* sp. (strain RCC299 / NOUM17); B4VSU9, *Coleofasciculus chthonoplastes* PCC7420; Q31Q05, *Synechococcus elongatus* PCC7942; Q5N472, *Synechococcus elongatus* PCC6301; B1XK11, *Synechococcus* sp. PCC7002.



Supplementary Figure 2

Functional analysis of Raf1 homologs.

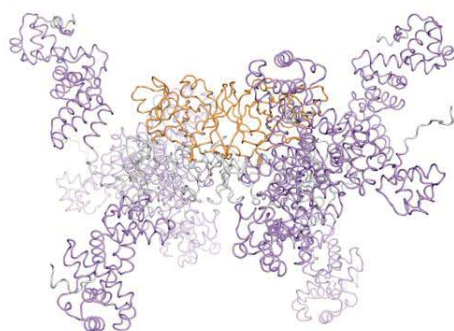
(a) Purified full-length Raf1 proteins and the respective α - and β -domains of Syn7942, Syn7002 and *A. thaliana*. AtRaf1.1/1.2 is a complex of AtRaf1.1 and AtRaf1.2 produced from a bicistronic plasmid. (b) Native-PAGE analysis of Rubisco reconstitution reactions as in Fig. 1b, containing the Raf1 proteins indicated. (c) Rubisco activities in reactions shown in (b) after supplementation with RbcS as described in Fig. 1d. Error bars, s.d. ($n = 3$ independent experiments). (d) Displacement of Raf1 from RbcL₈ by RbcS. Purified *S. elongatus* RbcL₈ was incubated with Syn7942-Raf1 as in Fig. 2a, followed by addition of RbcS (5μ M) and analysis by native-PAGE and immunoblotting with anti-RbcL (left) and anti-Raf1 (right). *S. elongatus* RbcL₈ and RbcL₈S₈ were used as standards. (e) Stoichiometry of RbcL and Raf1 in RbcL* complexes. RbcL* was excised from native-PAGE and separated by SDS-PAGE followed by Coomassie staining and quantitation by densitometry. Top, molar ratios of RbcL and Raf1 standards as quantified by extinction coefficients. Bottom, ratios of RbcL to Raf1 stain intensities are indicated as averages \pm S.D from four measurements. The molar ratio of RbcL to Raf1 in RbcL* is close 1:1. Shown is a representative Coomassie stained gel. (f) Dependence of Rubisco assembly on Raf1 concentration. Reconstitution reactions were performed as in Fig. 1e at increasing concentrations of Raf1 and the Rubisco activities obtained after 60 min are indicated as percentage of control. Error bars, s.d. ($n = 3$ independent experiments).



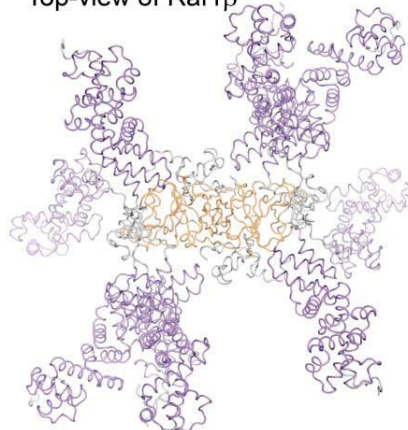
d

| | R_g (Å) | D_{max} (Å) |
|-------------------|-----------|---------------|
| AtRaf1.2 | 52 | 208 |
| AtRaf1.2 α | 24 | 78 |
| AtRaf1.2 β | 25 | 76 |
| Syn7942-Raf1 | 59 | 225 |

e Side-view of Raf1 β



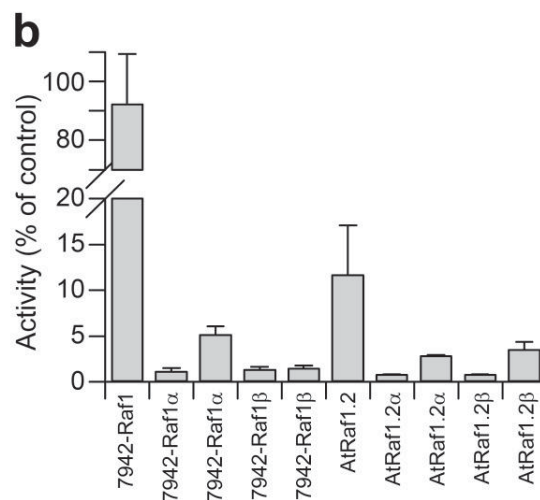
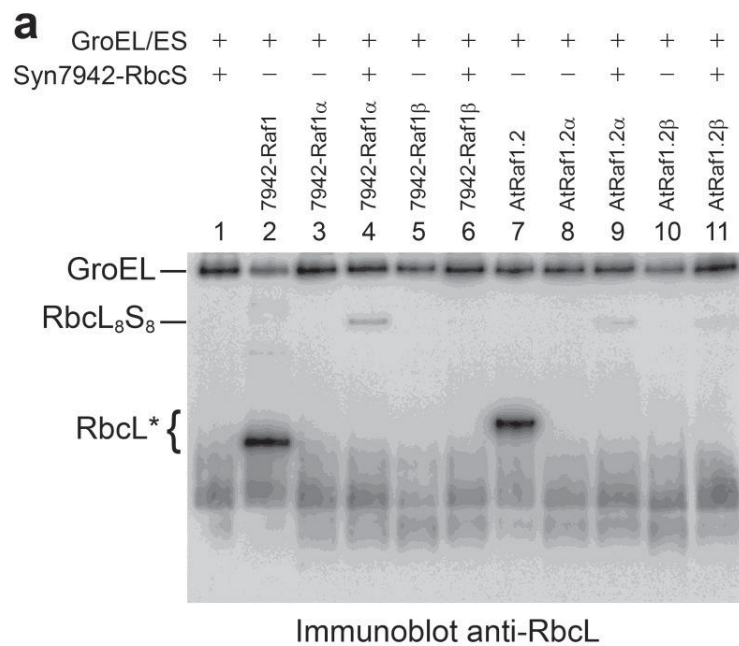
Top-view of Raf1 β



Supplementary Figure 3

SEC-MALS and SAXS analysis of Raf1 proteins.

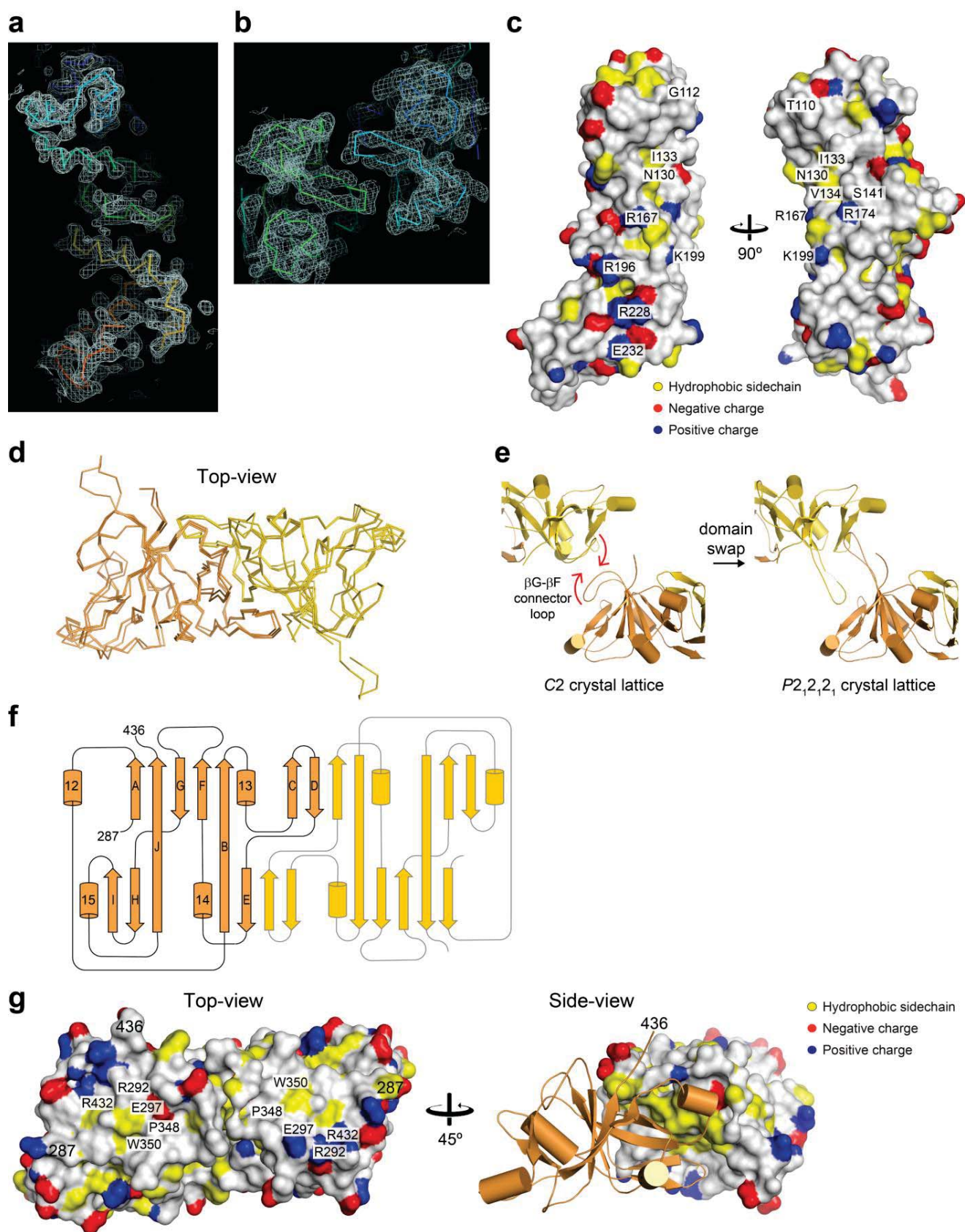
(a) SEC-MALS analysis of purified Raf1 domains from Syn7942, Syn7002 and *A. thaliana*. Data showing measurements for ~30 µg of the respective proteins. Horizontal lines across the peaks indicate molar mass and homogeneity of the sample. Calculated molar masses and hydrodynamic radii are indicated. (b) Representative X-ray scattering curves of AtRaf1.2 (red), AtRaf1.2 α (blue) and AtRaf1.2 β (green) and Syn7942-Raf1 (black). The curves represent background-corrected averages of ten measurements. The GNOM-fitted⁵¹ curves are overlaid. (c) Density distributions for AtRaf1.2 (red), AtRaf1.2 α (blue), AtRaf1.2 β (green) and Syn7942-Raf1 (black) calculated with GNOM. AtRaf1.2 α and AtRaf1.2 β appear rod-shaped and globular, respectively. The curves for AtRaf1.2 and Syn7942-Raf1 suggests flexibly linked domains. (d) Parameters from SAXS data analysis. Radii of gyration were determined using the Guinier approximation. Scattering curves were fitted with GNOM. (e) Ensemble model for the AtRaf1.2 dimer. Two perpendicular views are shown. The backbones are represented as coils. A subset of five models matching the experimental scattering curve (Chi value 3.978) was picked from a library of 10,000 conformations by a genetic algorithm implemented in the program EOM 2.0^{52,53}. The position of the dimeric β -domain (orange) was fixed at the dyad symmetry axis. The α -domains are represented in purple; the flexible termini and inter-domain linkers are shown in gray.



Supplementary Figure 4

Functional analysis of Raf1 homologs.

(a) Native-PAGE analysis of Rubisco reconstitution reactions as in Fig. 1b, containing purified full-length Raf1 and the α - and β -domains from Syn7942 and *A. thaliana*. RbcS was present when indicated. (b) Rubisco activities in reactions shown in (a) after supplementation with RbcS when absent, as described in Fig. 1d. Error bars, s.d. ($n = 3$ independent experiments).



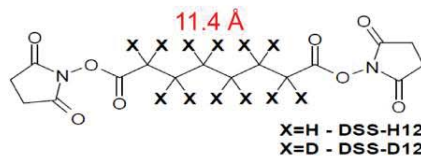
Supplementary Figure 5

Crystal structures of AtRaf1.2 domains.

(a,b) Experimental electron density maps for AtRaf1.2 α and AtRaf1.2 β . Representative regions are shown. The meshwork represents the isocontour surface at 1.0 σ level. The nominal resolutions of the AtRaf1.2 α Pt-SIRAS and AtRaf1.2 β Hg-SIRAS maps are 2.75 and 3.0 Å, respectively. Panel B shows a contact between two AtRaf1.2 β dimers. **(c)** Surface properties of AtRaf1.2 α . The same views as in Fig. 3c are shown. Positively and negatively charged groups are shown in blue and red, respectively. Yellow color signifies hydrophobic sidechains. **(d)** Superposition of three crystallographically independent copies of the AtRaf1.2 β dimer. The models are represented as C α traces. The orientation corresponds to the top-view in Fig. 3d. **(e)** Domain swapping in the $P2_12_12_1$ crystal lattice of AtRaf1.2 β . In this lattice the long β F- β G connecting loops reach across between adjacent dimers, making contacts to a hydrophobic pit. In the C2 crystal form, the hydrophobic residues at the loop apex fold back onto the respective hydrophobic area of the same chain, realizing an analogous intramolecular contact. Outside of the crystal lattice the conformation observed in the C2 crystal form should be strongly favored. **(f)** Topology of the secondary structure in the AtRaf1.2 β dimer. α -Helices and β -strands are represented by cylinders and arrows, using the same color scheme as in the main text. The monomer shown in orange differs from the second by insertion of helix 12. **(g)** Features of the AtRaf1.2 β dimer interface. On the left, the surface properties of the area facing the RbcL dimer are shown using the same representation as in panel c. On the right, one monomer is shown as backbone ribbon, the other in surface representation to reveal the AtRaf1.2 β dimer interface. Yellow color signifies hydrophobic sidechains.

a

Disuccinimidylsuberate (DSS):
bifunctional lysine-specific crosslinker
(1:1 isotopic mixture)



b

Syn6301-RbcL₈ Raf1

15 min, 25°C

DSS crosslinking 30 min, 25°C
Quench NH₄HCO₃

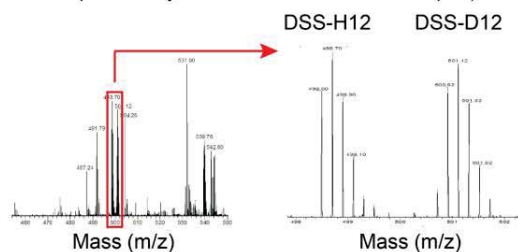
Separation on 4-12% SDS-PAGE

In-gel trypsin digestion

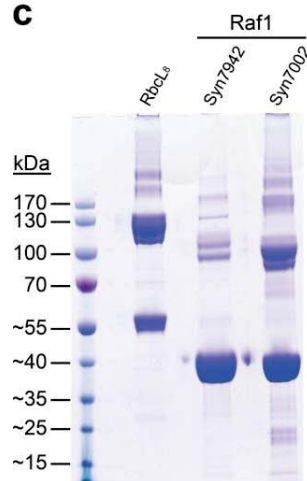
LC-MS/MS

Program StavroX to identify
crosslinked peptides
(b- and y-ions assignment)

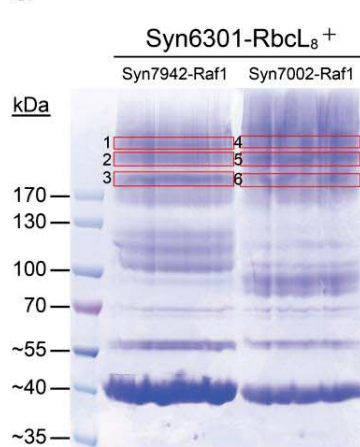
Manual validation of crosslinked peptides
(presence of two signals of equal intensity,
separated by mass difference of the isotopes)



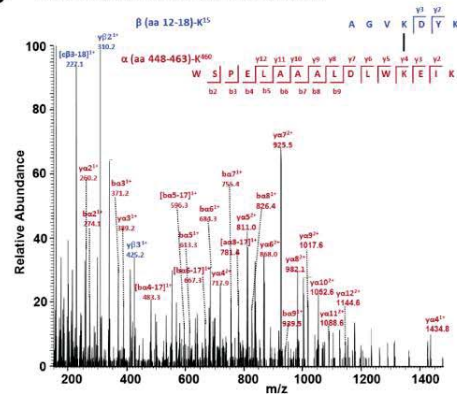
c



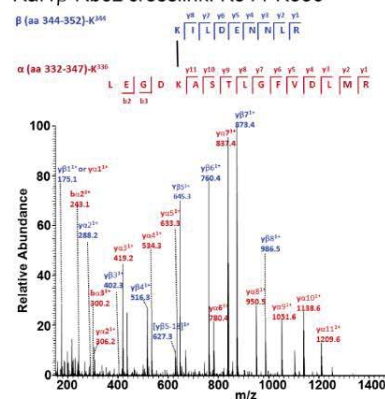
d



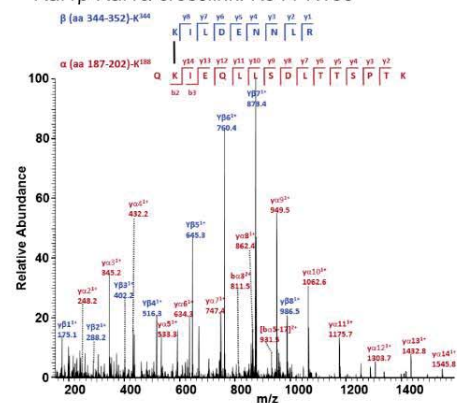
e RbcL-RbcL crosslink: K15-K460



Raf1β-RbcL crosslink: K344-K336



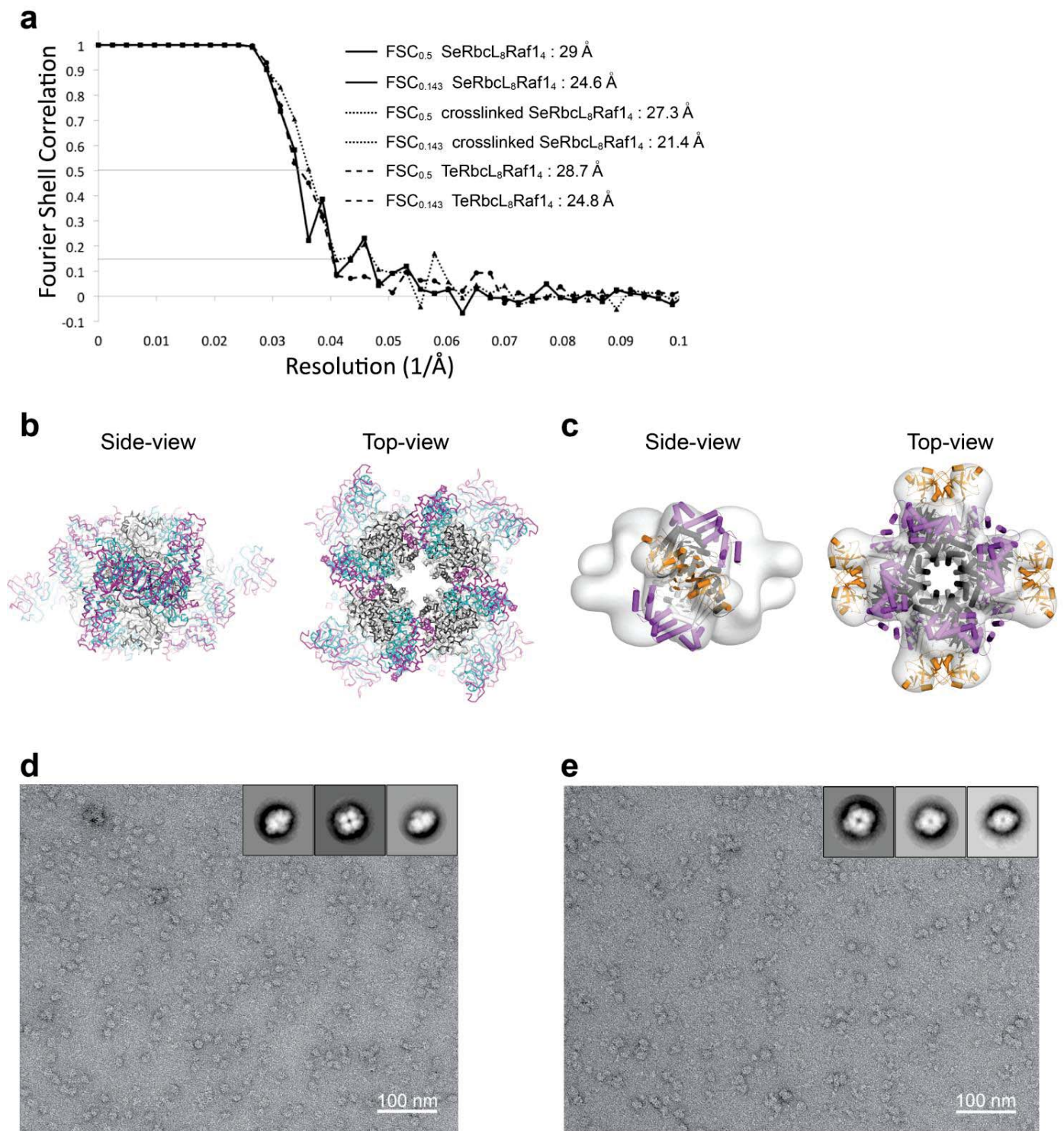
Raf1β-Raf1α crosslink: K344-K188



Supplementary Figure 6

Cross-linking coupled to mass spectrometry (CXMS).

(a) Structure of the bifunctional crosslinker disuccinimidylsuberate (DSS). The crosslinker is a 1:1 mixture of unlabeled (light; H_{12}) and deuterium labeled (heavy; D_{12}) compounds with a mass difference of 12.0753 Da. (b) Workflow for analysis of crosslinked protein bands marked and numbered in (c) by in-gel trypsin digestion, followed by LC-MS. (c) Crosslinking products of individual proteins *S. elongatus* RbcL₈, Syn7942-Raf1 and Syn7002-Raf1. The purified proteins (1.25 μ M RbcL₈ and 10 μ M of the respective Raf1 proteins) were incubated with H_{12}/D_{12} -DSS (2 mM) for 30 min at 25°C, followed by quenching of the crosslinking reaction with NH_4HCO_3 (150 mM) and analysis by SDS-PAGE. (d) Crosslinking products of RbcL₈ (1.25 μ M) with Syn7942-Raf1 or Syn7002-Raf1 (10 μ M each). Boxed areas were analyzed as in (b). (e) Representative MS/MS spectra for the crosslinks RbcL-RbcL (Lys15-Lys460), Raf1 β -RbcL (Lys344-Lys336) and Raf1 β -Raf1 α (Lys344-Lys188).

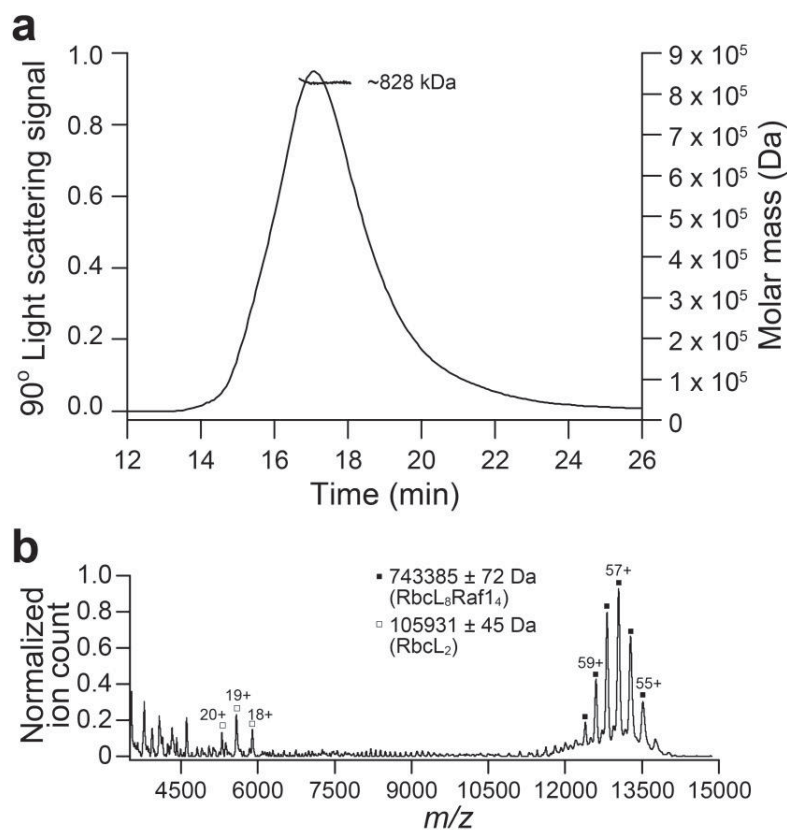


Supplementary Figure 7

Negative-stain EM analysis.

(a) Fourier Shell Correlation (FSC) curves of *S. elongatus* (Se) RbcL₈-Raf1₄, crosslinked SeRbcL₈-Raf1₄, and *T. elongatus* (Te)

RbcL₈–Raf1₄ as determined by gold standard FSC procedure in RELION-1.3. The resolution of the maps estimated by FSC with 0.5 and 0.143 correlation cut-off and no masking are given. **(b)** Comparison of the RbcL₈–Raf1₄ models derived from CXMS distance restraints (**Fig. 5g,h**) and EM reconstruction (**Fig. 6f,h,j**) (assuming the C-terminal 65 residues of RbcL are structured). The backbones are represented by C α traces. Raf1 and RbcL in the CXMS model are shown in magenta and white, respectively. Raf1 and RbcL in the EM reconstruction are shown in cyan and gray, respectively. **(c)** Rigid body domain fitting of SeRaf1 α - and β -domains and RbcL₈ missing the C-terminal 65 amino acids into the 3D reconstruction of SeRbcL₈–Raf1₄. RbcL subunits in gray and black; Raf1 α in purple and Raf1 β in orange. Side- and top-views are shown. **(d,e)** Negative stain electron micrograph of crosslinked SeRbcL₈–Raf1₄ (**d**) and of TeRbcL₈–Raf1₄ (**e**). Exemplary class averages of the respective complexes obtained from 2D classification in RELION-1.3 are shown in the insets.



Supplementary Figure 8

Characterization of the RbcL–Raf1 complex of the thermophilic cyanobacterium *T. elongatus*.

T. elongatus RbcL and Raf1 proteins were coexpressed in *E. coli* and purified as a high molecular weight complex. **(a)** SEC-MALS analysis of RbcL–Raf1 complex in solution (~40 μ g). The horizontal line across the peak indicates the calculated molar mass. Note that the sample contained a small amount of aggregated protein which leads to a higher average molar mass (~828 kDa) than expected for the RbcL₈–Raf1₄ complex (~740 kDa). **(b)** nano-ESI native MS spectra of RbcL–Raf1 complex (~8 μ M). Symbols indicate the charge state distributions with the charge states shown for some peaks; the calculated mass around the m/z values of the respective protein complexes is indicated. S.D. refers to the accuracy of mass values calculated from the different m/z peaks. The theoretical masses for RbcL₈–Raf1₄ and RbcL₂ are 741628.8 Da and 106265.4 Da, respectively.

SUPPLEMENTARY TABLES

Supplementary Table 1. PCR Primer Sequences for Cloning of Raf1 Constructs. The SacI (GAGCTC), SacII (CCGCGG) and NotI (GCGGCCGC) restriction sites are underlined, and the start codons are italicized.

| Plasmid | | Oligos |
|----------------------------------|-------------------|--|
| <i>Syn7942raf1</i> | Syn7942Raf1: 5' | <u>CCGCGGTGGTATGCGTGAGTTCACCCCCAC</u> |
| | Syn7942Raf1: 3' | <u>GAGCTCCTATTCTTCAA</u> ACTGCCAAGGTG |
| <i>Syn7002raf1</i> | Syn7002Raf1: 5' | <u>CCGCGGTGGTATGATTGGACAACCGCAATCTCC</u> |
| | Syn7002Raf1: 3' | <u>GAGCTCTTAATCGTCCATT</u> TGCCAGGGTTCC |
| <i>Atraf1.1</i> (52-434) | AtRaf1.1: 5' | <u>CCGCGGTGGTCAACA</u> ACTCTATCAACCATTCCG |
| | AtRaf1.1: 3' | <u>GAGCTCCTAATCCCAATCTTCATCATT</u> TATC |
| <i>Atraf1.2</i> (62-449) | AtRaf1.2: 5' | <u>CCGCGGTGGTCAACAGCTCTACCAACCGTTCCG</u> |
| | AtRaf1.2: 3' | <u>GAGCTCTCAGTCCCAGTTCTGATGACTTG</u> |
| <i>Atraf1.1/1.2</i> | AtRaf1.1/1.2: 5' | <u>GAGCTCAAGCTTAGATCCGGCTGCTAACA</u> AAGCC CGTAAGAAGGAGATATAACCATGCAACAGCTCTACC AACCGTTC |
| | AtRaf1.1/1.2: 3' | <u>GCGGCCGCTCAGTCCCAGTTCTGATGACTTGTCTG</u> C |
| <i>Syn7942raf1α</i> (1-192) | Syn7942Raf1α: 5' | <u>CCGCGGTGGTATGCGTGAGTTCACC</u> |
| | Syn7942Raf1α: 3' | <u>GAGCTCTTAATCTAGCAGCAATGCTTCGATGAGG</u> |
| <i>Syn7942raf1β</i> (194-356) | Syn7942Raf1β: 5' | <u>CCGCGGTGGTCCGATGCTGCCGATCTATCGACTG</u> |
| | Syn7942Raf1β: 3' | <u>GAGCTCCTATTCTTCAA</u> ACTGCCAAGGTGTTG |
| <i>Syn7002raf1α</i> (1-195) | Syn7002Raf1 α: 5' | <u>CCGCGGTGGTATGATTGGACAACCGCAATCTCC</u> |
| | Syn7002Raf1 α: 3' | <u>GAGCTCTTAGTCACTGAGGAGTTGCTCGAT</u> |
| <i>Syn7002raf1β</i> (197-359) | Syn7002Raf1 β: 5' | <u>CCGCGGTGGTCCGCTCCTCCCCCTCTACCGTTAC</u> |
| | Syn7002Raf1 β: 3' | <u>GAGCTCTTAATCGTCCATT</u> TGCCAGGG |
| <i>Atraf1.2α</i> (62-274) | AtRaf1.2α: 5' | <u>CCGCGGTGGTCAACAGCTCTACCAACCGTTCCG</u> |
| | AtRaf1.2α: 3' | <u>GAGCTCTCACTTCTCCTTCTCTGCTTCCTTGTC</u> |
| <i>Atraf1.2β</i> (281-449) | AtRaf1.2β: 5' | <u>CCGCGGTGGTGAGGAAGTCAAAGCTATTCGGATTC</u> C |
| | AtRaf1.2β: 3' | <u>GAGCTCTCAGTCCCAGTTCTGATGACTTG</u> |

SUPPLEMENTARY NOTE

Supplementary Data Set 1 DSS crosslinks identified in the complex of *S. elongatus* RbcL₈ and Raf1 **(a)** *S. elongatus* RbcL₈ and Syn7942-Raf1 complex. Protein band refers to the band number on SDS-PAGE in Supplementary Fig. 6d; A, experiment 1; B, experiment 2. m/z, mass over charge. The crosslinked peptides, as well as their detected and calculated mass, and the positions of the crosslinked residues are indicated. m, denotes oxidized methionine. Note that the N-terminal amino acid of a protein can be crosslinked with DSS. See Online Methods and Supplementary Fig. 6 for details. **(b)** Consolidated list of crosslink sites from Supplementary Data Set 1a. Domain categories: N, N-terminal residues 1–11 in Syn7942-Raf1 and residues 1–147 in *S. elongatus* RbcL; TIM, TIM-barrel residues 148–410 in *S. elongatus* RbcL; C, C-terminal residues 411–472 in *S. elongatus* RbcL; α , α -domain residues 12–192 in Syn7942-Raf1; β , β -domain residues 202–342 in Syn7942-Raf1; linker, residues 193–201 in Syn7942-Raf1. Crosslink distances between RbcL residues indicate the intra- or intermolecular distance in RbcL₂ in context of the RbcL₈ core complex. Crosslink distances between RbcL and Raf1 residues refer to the distances of the respective residues in the structural model derived from crosslinking data (assuming the C-terminal 65 residues of RbcL are structured) or EM reconstruction (assuming the C-terminal 65 residues of the RbcL subunit are either structured or disordered). Distances of Raf1 to the adjacent RbcL and the next closest RbcL are reported. Crosslink distances between Raf1 residues indicate the intra- or intermolecular distance in the Raf1 dimer in the structural models derived from crosslinking data (assuming the C-terminal 65 residues of RbcL are structured) or EM reconstruction (assuming the C-terminal 65 residues of the RbcL subunit are either structured or disordered). ¹, RbcL residues in the flexible N- or C-terminal sequence precluding distance assignment. ², Raf1 residues in the flexible N-terminal sequence and interdomain linker precluding distance assignment. *, denotes a crosslink between Raf1 α -domains that cannot occur in the context of the structural models of the RbcL₈Raf1₄ complex and must be due to the dynamic nature of the interaction. **(c)** *S. elongatus* RbcL₈ and Syn7002-Raf1 complex. Protein band refers to the band number on SDS-PAGE in Supplementary Fig. 6d; A, experiment 1; B, experiment 2. m/z, mass over charge. The crosslinked peptides, as well as their detected and calculated mass, and the positions of the crosslinked residues are indicated. m, denotes oxidized methionine. Note that the N-terminal amino acid of a protein can be crosslinked with DSS. See Online Methods and Supplementary Fig. 6 for details. **(d)** Consolidated list of crosslink sites from Supplementary Data Set 1c. Domain categories: N, N-terminal residues 1–14 in Syn7002-Raf1 and residues 1–147 in *S. elongatus* RbcL; TIM, TIM-barrel residues 148–410 in *S. elongatus* RbcL; C, C-terminal residues 411–472 in *S. elongatus* RbcL; α , α -domain residues 15–195 in Syn7002-Raf1; β , β -domain residues 205–345 in Syn7002-Raf1; linker, residues 196–204 in Syn7002-Raf1. Crosslink distances between RbcL residues indicate the intra- or intermolecular distance in RbcL₂ in context of the RbcL₈ core complex. Intermolecular distances of RbcL to the next closest RbcL of a RbcL₂ are also reported. Crosslink distances between RbcL and Raf1 residues refer to the distances of the respective residues in the structural model derived from crosslinking data (assuming the C-terminal 65 residues of RbcL are structured). Distances of Raf1 to the adjacent RbcL and the next closest RbcL are reported. Crosslink distances between Raf1 residues indicate the intra- or intermolecular distance in the Raf1 dimer in the structural model derived from crosslinking data (assuming the C-terminal 65 residues of RbcL are structured). ¹, RbcL residues in the flexible N- or C-terminal sequence precluding distance assignment. ², Raf1

residues in the flexible N-terminal sequence and interdomain linker precluding distance assignment. **(e)** Reconstitution of SeRbcL in the presence Syn7942-Raf1. Protein band refers to the band number on SDS-PAGE. A, experiment 1; B, experiment 2. m/z, mass over charge. The crosslinked peptides, as well as their detected and calculated mass, and the positions of the crosslinked residues are indicated. m, denotes oxidized methionine. Note that the N-terminal amino acid of a protein can be crosslinked with DSS. See Online Methods and Supplementary Fig. 6 for details. **(f)** Consolidated list of crosslink sites from Supplementary Data Set 1e. Domain categories: N, N-terminal residues 1–14 in Syn7002-Raf1 and residues 1–147 in *S. elongatus* RbcL; TIM, TIM-barrel residues 148–410 in *S. elongatus* RbcL; α , α -domain residues 15–195 in Syn7002-Raf1; β , β -domain residues 205–345 in Syn7002-Raf1; linker, residues 196–204 in Syn7002-Raf1. Crosslink distances between RbcL residues indicate the intra- or intermolecular distance in RbcL₂ in context of the RbcL₈ core complex. Crosslink distances between RbcL and Raf1 residues refer to the distances of the respective residues in the structural model derived from crosslinking data (assuming the C-terminal 65 residues of RbcL are structured). Crosslink distances between Raf1 residues indicate the intramolecular distance in the Raf1 dimer in the structural model derived from crosslinking data (assuming the C-terminal 65 residues of RbcL are structured). ², Raf1 residues in the flexible interdomain linker precluding distance assignment.

2.2 Structural Analysis of the Rubisco-Assembly Chaperone RbcX-II from *Chlamydomonas reinhardtii*

RESEARCH ARTICLE

Structural Analysis of the Rubisco-Assembly Chaperone RbcX-II from *Chlamydomonas reinhardtii*

Andreas Bracher^{*†}, Thomas Hauser[‡], Cuimin Liu[‡], F. Ulrich Hartl, Manajit Hayer-Hartl^{*}

Department of Cellular Biochemistry, Max-Planck-Institute of Biochemistry, Martinsried, Germany

[‡] These authors contributed equally to this work.

[‡] Current address: State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

* bracher@biochem.mpg.de (AB); mhartl@biochem.mpg.de (MH-H)



OPEN ACCESS

Citation: Bracher A, Hauser T, Liu C, Hartl FU, Hayer-Hartl M (2015) Structural Analysis of the Rubisco-Assembly Chaperone RbcX-II from *Chlamydomonas reinhardtii*. PLoS ONE 10(8): e0135448. doi:10.1371/journal.pone.0135448

Editor: Yong-Bin Yan, Tsinghua University, CHINA

Received: June 3, 2015

Accepted: July 22, 2015

Published: August 25, 2015

Copyright: © 2015 Bracher et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data for this study are accessible through the RCSB Protein Data Bank at (<http://www.rcsb.org/pdb/explore/explore.do?structureId=5BS1>) and (<http://www.rcsb.org/pdb/explore/explore.do?structureId=5BS2>).

Funding: The authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

The most prevalent form of the Rubisco enzyme is a complex of eight catalytic large subunits (RbcL) and eight regulatory small subunits (RbcS). Rubisco biogenesis depends on the assistance by specific molecular chaperones. The assembly chaperone RbcX stabilizes the RbcL subunits after folding by chaperonin and mediates their assembly to the RbcL₈ core complex, from which RbcX is displaced by RbcS to form active holoenzyme. Two isoforms of RbcX are found in eukaryotes, RbcX-I, which is more closely related to cyanobacterial RbcX, and the more distant RbcX-II. The green algae *Chlamydomonas reinhardtii* contains only RbcX-II isoforms, CrRbcX-IIa and CrRbcX-IIb. Here we solved the crystal structure of CrRbcX-IIa and show that it forms an arc-shaped dimer with a central hydrophobic cleft for binding the C-terminal sequence of RbcL. Like other RbcX proteins, CrRbcX-IIa supports the assembly of cyanobacterial Rubisco in vitro, albeit with reduced activity relative to cyanobacterial RbcX-I. Structural analysis of a fusion protein of CrRbcX-IIa and the C-terminal peptide of RbcL suggests that the peptide binding mode of RbcX-II may differ from that of cyanobacterial RbcX. RbcX homologs appear to have adapted to their cognate Rubisco clients as a result of co-evolution.

Introduction

Life on earth depends on fixation of atmospheric CO₂ into organic compounds by bacteria, algae and plants. The key enzyme for this process ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) catalyzes the carboxylation of the five-carbon sugar ribulose-1,5-bisphosphate (RuBP) which is converted into two molecules of 3-phosphoglycerate. The other enzymes of the Calvin–Benson–Bassham cycle subsequently use reduction equivalents and ATP produced in the light reaction of photosynthesis to regenerate RuBP and produce triose phosphate to fuel anabolic pathways. The most prevalent form of Rubisco (form I) consists of a complex of eight catalytic large subunits (RbcL), forming a D₄-symmetric core, and eight

regulatory small subunits (RbcS), capping the RbcL₈ complex at both ends [1]. RbcL sequences exhibit remarkable conservation across phyla. Nevertheless, based on sequence diversity of the RbcL subunits, four subgroups of form I Rubisco, IA-ID, can be distinguished [2, 3]. The economically most important form IB is found in so-called green organisms: cyanobacteria, green algae and plants.

While there is a plethora of data on Rubisco structure, function and catalysis [1, 4], the pathways of subunit folding and oligomeric assembly are only beginning to emerge [5]. In green algae and plants, the RbcL subunits are chloroplast encoded and synthesized in the chloroplast stroma, the site of carbon fixation. In contrast, the RbcS subunits are nuclear-encoded, translated in the cytosol and imported into chloroplasts [6]. Newly-synthesized RbcL subunits associate with the chloroplast chaperonin Cpn60 [7], the homolog of bacterial GroEL, initially suggesting that the chaperonin mediates Rubisco assembly [8].

Recent reconstitution of cyanobacterial form I Rubisco in vitro demonstrated that the chaperonin mediates RbcL folding, while assembly of the RbcL₈ core complex requires the additional factor RbcX [9, 10]. Co-expression of RbcX was also required for the recombinant production of the Rubisco from the cyanobacterial species *Synechococcus* sp. PCC7002 (Syn7002) and increased the efficiency of functional expression of *Synechococcus elongatus* PCC6301 (Syn6301) Rubisco [11, 12]. In most cyanobacteria, the gene for RbcX is located between the *rbcL* and *rbcS* genes within a single operon [13]. Mutation or deletion of *rbcX* was found to reduce the level of functional Rubisco in PCC7002 and *Synechococcus elongatus* PCC7942 [11, 14]. RbcX is highly conserved in all prokaryotes and eukaryotes containing form IB Rubisco [15]. Structural analysis showed that RbcX is a dimeric, α -helical protein of ~15 kDa subunits [12, 16–18]. The dimer structure has a central hydrophobic cleft which serves as binding site for the C-terminal sequence motif EIKFEF(E/D) in RbcL sequences [12, 15]. The peptide binds in an extended conformation with the Phe sidechains reaching into hydrophobic pockets [10, 12]. In addition, the boomerang-shaped RbcX dimer has conserved residues at the corners that mediate interactions with the adjacent RbcL subunit [10, 12]. Thus, RbcX binding clamps the RbcL anti-parallel dimer together and facilitates the assembly of the RbcL₈ core complex. The RbcL-RbcX interaction is dynamic, allowing the displacement of RbcX from RbcL₈ complexes by RbcS to form the holoenzyme. RbcX therefore functions as a Rubisco assembly chaperone.

Many eukaryotes have two RbcX homologs, one that closely resembles the cyanobacterial ortholog, RbcX-I, and a more distantly related homolog, RbcX-II [12]. The RbcX-I and RbcX-II from *Arabidopsis thaliana* have been characterized and crystallized, named AtRbcX2 and AtRbcX1, respectively, in these studies [18, 19]. The green algae *Chlamydomonas reinhardtii* contains two RbcX-II sequences (CrRbcX-IIa and CrRbcX-IIb, orthologs of AtRbcX-II) and no RbcX-I ortholog. Here we biochemically and structurally characterize CrRbcX-IIa. The crystal structures of CrRbcX-IIa alone and in complex with the C-terminal peptide of RbcL show that CrRbcX-IIa shares the structural topology with cyanobacterial and plant RbcX homologs. However, the RbcL peptide bound to CrRbcX-IIa only occupies part of the central hydrophobic cleft of the RbcX dimer, in contrast to the structure of the cyanobacterial RbcX-peptide complex. Nevertheless, we find that CrRbcX-IIa supports the assembly of cyanobacterial Rubisco, although with reduced efficiency compared to cyanobacterial RbcX-I.

Materials and Methods

Plasmids and Proteins

Open reading frames for CrRbcX-IIa were amplified by PCR from *Chlamydomonas reinhardtii* cDNA [20], and cloned between the SacII and SacI restriction sites of the pHue plasmid [21],

resulting in the following constructs: *pHueCrRbcX-IIa(33–189)*; *pHueCrRbcX-IIa(34–156)*; *pHueCrRbcX-IIa(34–189)*; *pHueCrRbcL(462–474)-RbcX-IIa(37–156)*. The cleavage site for the chloroplast transit peptide of CrRbcX-IIa was predicted based on homology with AtRbcX-II (see Fig 1). The Quik-Change protocol (Stratagene) was used to produce the mutant *pHueCrRbcX-IIa(33–189)(R118A)*. All plasmid inserts were verified by DNA sequencing.

S. elongatus PCC6301 RbcL_{8S8}, RbcL₈, RbcS, RbcL, GroEL and GroES were purified as previously described [9, 12, 22]. Rabbit antibody against *S. elongatus* PCC6301 RbcL was produced using standard procedures.

Expression and Purification of CrRbcX-IIa

RbcX proteins were expressed as N-terminal His₆-ubiquitin (His₆-Ub) fusion proteins in *E. coli* BL21 (DE3) cells from *pHue* expression plasmids. Cells were grown to an OD₆₀₀ of 0.5 at 37°C in LB medium followed by induction for 16 h with 0.5 mM isopropyl-D-thiogalactoside (IPTG) at 23°C. Cells were lysed in 50 mM Tris-HCl pH 8.0, 20 mM NaCl, 1 mM EDTA, 0.5 mg/ml lysozyme and 5 mM phenylmethylsulfonyl fluoride (PMSF) for 30 min on ice, followed by ultrasonication (Misonix Sonicator 3000). The supernatant obtained after high-speed centrifugation (48 000 x g, 40 min, 4°C) was applied to a Ni-IMAC column (GE Biotech) to capture the His₆-Ub protein, followed by overnight cleavage of the His₆-Ub moiety at 23°C using the deubiquitinating enzyme Usp2 [21]. All subsequent steps were performed at 4°C. The supernatant was dialyzed against buffer A (20 mM Tris-HCl pH 8.0, 50 mM NaCl) and applied to a pre-equilibrated MonoQ column (GE Biotech). Proteins were eluted with a linear salt gradient from 50 mM to 1 M NaCl. Fractions containing RbcX were combined and concentrated, 5% glycerol was added, followed by flash-freezing in liquid nitrogen and storage at -80°C.

RbcX for crystallographic studies was purified further by Superdex200 (GE Biotech) size exclusion chromatography in buffer A. Protein concentration was determined spectrophotometrically at 280 nm using calculated extinction coefficients.

For selenomethionine (SeMet) labeling by the catabolite repression method [23], the bacteria were grown to mid-log phase at 37°C in M9 medium containing 100 mg L⁻¹ ampicillin. Methionine biosynthesis repression was induced by addition of amino acids as follows: 125 mg L⁻¹ L-Lys, 100 mg L⁻¹ L-Phe, 100 mg L⁻¹ L-Tyr, 50 mg L⁻¹ L-Ile, 50 mg L⁻¹ L-Leu, 50 mg L⁻¹ L-Val and 60 mg L⁻¹ L-Se-Met. 15 min later the temperature was reduced to 23°C and protein synthesis was induced with 0.5 mM IPTG for 20 h. Cells were harvested and re-suspended in lysis buffer (50 mM Na-phosphate pH 9.0, 300 mM NaCl, 10 mM imidazole and 1 mM β-mercaptoethanol) containing Complete protease (Roche Biotech) inhibitor cocktail. The cells were disrupted by ultrasonication and SeMet-labeled His₆-Ub RbcX was purified essentially as described above. The protein solution was dialyzed against buffer A containing 1 mM β-mercaptoethanol (β-ME) and applied to a pre-equilibrated MonoQ column. Proteins were eluted with a linear salt gradient from 50 to 400 mM NaCl. Fractions containing SeMet-labeled CrRbcX-IIa(34–156) were subsequently dialyzed against buffer A/β-ME and concentrated. After flash-freezing in liquid N₂, the protein was stored at -80°C.

Native Mass Spectrometry (Native-MS)

Purified CrRbcX-IIa(33–189); CrRbcX-IIa(34–156) and CrRbcX-IIa(34–189) (15 μM monomer each) were buffer-exchanged into 100 mM ammonium acetate pH 8.5 (Fluka, Sigma), using Micro Bio-Spin 6 chromatography columns (BioRad). Native-MS analyses were performed in positive ion mode on an electrospray ionization quadrupole time-of-flight (ESI-TOF) hybrid mass spectrometer (Synapt G2-Si, Waters Corp., Manchester, UK) equipped with a Z-spray nano-ESI source. The instrument was mass calibrated using a solution of 30 mg ml⁻¹

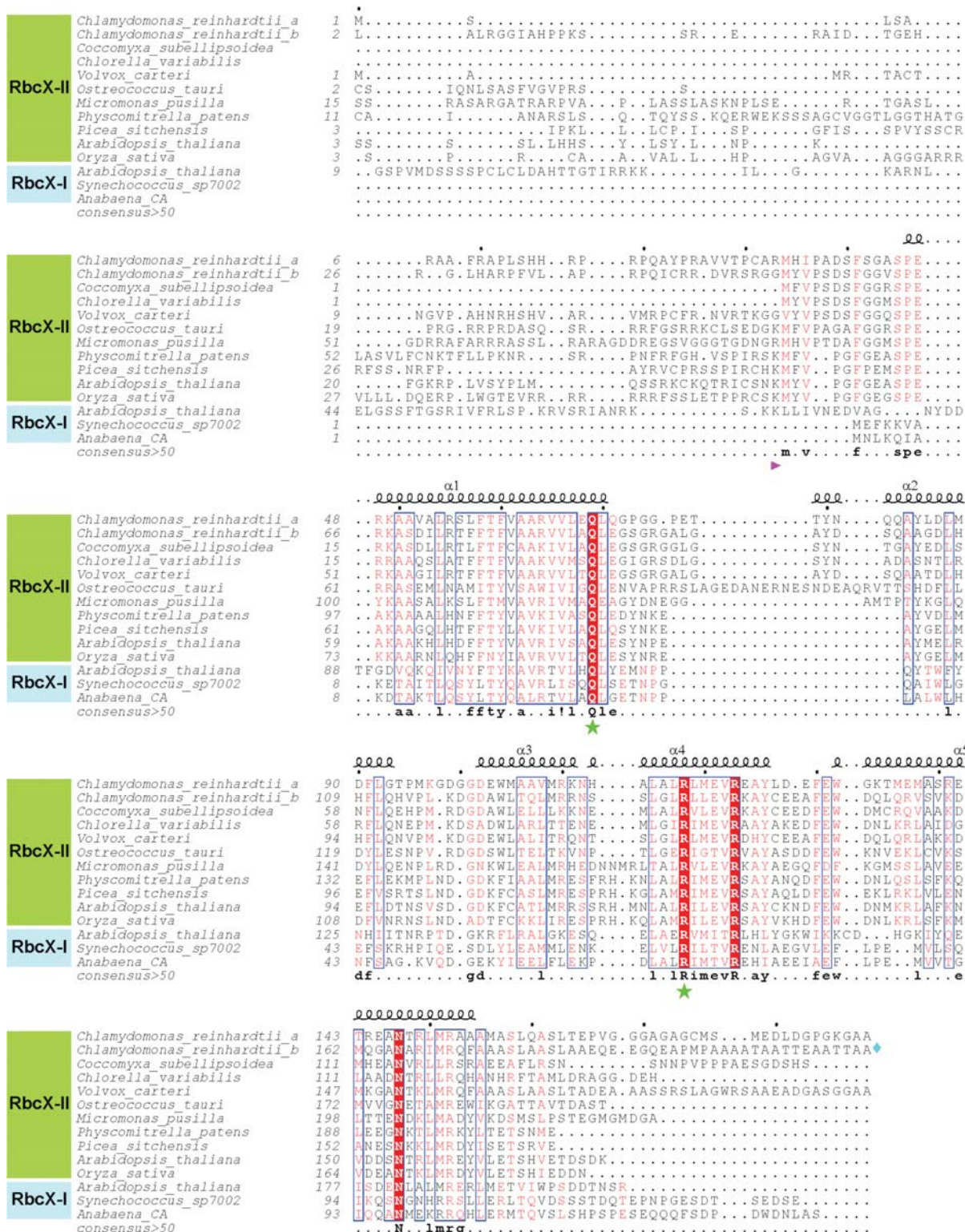


Fig 1. Sequence alignment of RbcX-II from green algae. Amino acid sequences of selected RbcX-II homologs from green algae, mosses and plants were aligned using Clustal-Ω. Note that for the green algae *Coccomyxa subellipsoidea*, *Chlorella variabilis*, *Volvox carteri*, *Ostreococcus tauri* and *Micromonas pusilla* only one RbcX-II sequence is shown. For comparison, RbcX-I from *A. thaliana*, *Synechococcus* sp. PCC7002 and *Anabaena* sp. CA are also aligned. All sequence numbering is based on the open reading frames. Secondary structure elements are indicated above the sequences. In the alignment, similar

residues are shown in red and identical residues in white using bold lettering on red background. Blue frames indicate homologous regions. The consensus sequence is shown at the bottom. The forward arrow designates the beginning of the mature RbcX-II proteins. The diamond symbol at the end of the CrRbcX-IIb sequence indicates that the sequence continues with 130 amino acids not displayed. Asterisks denote residues known to be essential for RbcX function.

doi:10.1371/journal.pone.0135448.g001

cesium iodide dissolved in 1:1 acetonitrile:water. Gold-plated 10 μ m nano-ESI pipettes (Mascom, Bremen) were used as capillaries. Optimized capillary and sample cone voltages were 1–1.3 kV and 100–150 V, respectively.

Rubisco Reconstitution

GroEL/ES-mediated RbcL folding was performed as in Liu et al. (2010) with modifications. Denatured *S. elongatus* PCC6301 RbcL was diluted 200-fold from 6 M GuHCl (final RbcL concentration 0.5 μ M) into ice-cold buffer B (20 mM MOPS-KOH pH 7.5, 100 mM KCl, 5 mM MgOAc₂) containing GroEL (1 μ M oligomer), 1 mg/ml BSA and 5 mM DTT. The reaction was incubated on ice for 60 min, followed by centrifugation to remove any aggregated RbcL. GroES (2 μ M oligomer), RbcX (2 μ M AnaCa-RbcX or 30 μ M CrRbcX dimer) and *S. elongatus* PCC6301 RbcS (5 μ M) were added to the supernatant containing GroEL bound Syn6301-RbcL as indicated in Figure legend. Reconstitution was initiated by addition of 4 mM ATP at 25°C. Reactions were stopped by addition of apyrase (Sigma) to a final concentration of 0.25 U/ μ l to inhibit GroEL/ES activity.

For measurement of Rubisco enzymatic activity at 25°C, the reaction was supplemented with Syn6301-RbcS (5 μ M) and C-terminal RbcL peptide (KEIKFEFETMD) of *S. elongatus* PCC6301 (200 μ M) when indicated, and assembly of RbcL₈S₈ allowed to proceed for 15 min before enzymatic assay. Rubisco carboxylation activity was determined after incubation for 10 min in 50 mM Tris-HCl pH 8.0, 10 mM MgCl₂, 30 mM NaH¹⁴CO₃ (25 Bq/nmol) and the amount of carbon fixed was quantified [24]. Activities are expressed as percent of purified Syn6301-RbcL₈ (~0.06 μ M oligomer) standard supplemented with RbcS (5 μ M).

Crystallization

Crystals of CrRbcX-IIa(34–156) were grown using the hanging drop vapor diffusion method at 20°C by mixing 1 μ l protein sample at 6 g L⁻¹ and 1 μ l reservoir solution. Crystals of SeMet-labeled CrRbcX-IIa(34–156) resembling shields were obtained with a reservoir solution containing 5% PEG-3350/0.2 M MgCl₂/50 mM Tris-HCl pH 8.0. For cryo-protection, the crystals were transferred into mother liquor containing 15% PEG-3350/0.2 M MgCl₂/50 mM Tris-HCl pH 8.0, followed by stepwise increase to 20% glycerol content and flash-freezing in liquid nitrogen.

Crystals of the CrRbcL(462–473)-RbcX-IIa(37–156) fusion protein were grown by the hanging drop vapor diffusion method at 20°C using 0.1M Tris-HCl pH 8.5, 25% PEG2000 monomethyl ether as precipitant.

Structure Solution and Refinement

The diffraction data were collected at beamline X10SA of the Swiss Light Source (SLS) in Villigen, Switzerland. Diffraction data were integrated and scaled with XDS [25]. Pointless [26], Scala [27] and Truncate [28] were used to convert the data to CCP4 format, as implemented in the CCP4i interface [29].

The structure of CrRbcX-IIa(34–156) was solved by Se-SAD using crystals from SeMet-labeled protein at 2.0 Å resolution. 36 selenium sites were found by direct methods using

SHELXD as implemented in HKL2MAP [30, 31]. SHELXE was used for density modification and auto-building of a poly-alanine model. The resulting map was readily interpretable and the sequence was docked using Coot [32]. The final model was created by iterative Coot model building and Refmac5 refinement cycles [33]. The structure of the fusion protein CrRbcL(462–474)-RbcX-IIa(37–156) was solved by molecular replacement using Molrep [34], and the models modified and refined as above. Residues facing solvent channels with disordered side chains were modeled as alanines. Coordinates were aligned with Lsqkab and Lsqman [35]. Figures were generated with the program PyMOL [36] and ESPript [37]. Coordinates and structure factor amplitudes were deposited to Protein Data Bank under accession codes 5BS1 and 5BS2.

Results

Structural Analysis of *Chlamydomonas reinhardtii* RbcX

The genome of *C. reinhardtii* contains no RbcX-I, but instead has two RbcX-II genes, *g688.t1* (locus Cre01.g030350) and *g7885.t1* (locus Cre07.g339000). We refer to the gene products as CrRbcX-IIa and CrRbcX-IIb, respectively. Note that in the most recent genome annotation CrRbcX-IIa would start at amino-acid residue 34 and lacks the sequence encoding the transit peptide. CrRbcX-IIb, on the other hand, has a putative transit peptide but the annotated gene codes for a protein twice the length of other RbcX homologs (~290 residues) with only the first ~160 amino acids displaying homology to RbcX proteins (Fig 1). The additional sequence in CrRbcX-IIb probably represents an intron, and thus the sequence for CrRbcX-IIb is apparently incorrectly annotated. We focused our analysis on CrRbcX-IIa, which was previously annotated with a putative transit peptide. Based on sequence alignment with the mature form of *A. thaliana* RbcX-II (also known as AtRbcX1), which begins with Lys46 [19], we cloned CrRbcX-IIa starting at Arg33 (Fig 1), generating a protein of ~17 kDa. CrRbcX-IIa(33–189) was recombinantly expressed and purified from the soluble fraction. Analysis by native-MS showed that CrRbcX-IIa is a dimer in solution, as expected (Fig 2A).

Full-length CrRbcX-IIa failed to crystallize. A stable fragment comprising residues 34–156 lacking the flexible C-terminal 33 residues was produced by subtilisin treatment, as determined by mass spectrometry (MS). An unstructured C-terminus was also found to be present in the cyanobacterial Syn7002-RbcX and was not required for function in Rubisco assembly [12]. We recombinantly expressed and purified the truncated CrRbcX-IIa(34–156) protein for further structural analysis. The structure of the selenomethionine (SeMet)-labeled CrRbcX-IIa(34–156) protein was solved by selenium-single-wavelength anomalous dispersion (Se-SAD) at 2.0 Å resolution. The experimental electron density was readily interpretable (Fig 3A). The structural model was built against data to 1.6 Å resolution and refined to final R and R_{free} values of 0.177 and 0.206, respectively (see Table 1 for data collection and refinement statistics). The asymmetric unit of the monoclinic unit cell contains four copies of CrRbcX-IIa(34–156) in a two-fold symmetric topology (Fig 3B). Each chain consists of a succession of five α -helices. In three of the subunits the insertion after helix α 1, residues 73–77, is disordered. This insertion is typical for RbcX-II sequences from green algae (Fig 1). Apart from the N-terminal 10 residues (see below), the backbones of the CrRbcX-IIa(34–156) subunits are closely similar (r. m.s.d. of C α positions of 0.267 to 0.577 Å). The subunits form arch-shaped, two-fold symmetric dimers with a hydrophobic cleft in the center (Fig 4A), similar to other known RbcX structures [12, 17, 18]. In each subunit helices α 1– α 4 form a four-helix bundle, which associates with helix α 5 of the opposing subunit in the dimer (Fig 4A). The N-terminal sequence of one subunit binds into the central cleft, with residues Met34 and Ile36 reaching into hydrophobic pockets located between the anti-parallel helices α 1 and α 1' at the bottom of the cleft (Fig 4B). The N-terminal ammonium group of Met34 engages in a tight salt bridge (lengths 2.53 and

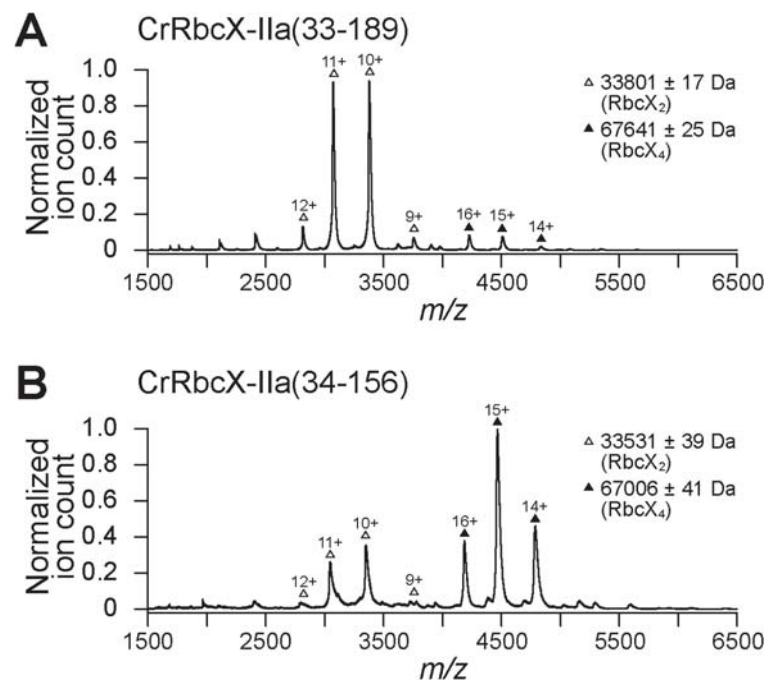


Fig 2. Oligomeric state of CrRbcX-IIa analyzed by native-MS. Nano-ESI native-MS spectra of CrRbcX-IIa (33–189) (A) and CrRbcX-IIa(34–156) (B). Symbols indicate the charge state distributions with the charge states shown for some peaks; the calculated mass around the m/z values of the respective protein complexes is reported. The accuracy of mass values calculated from the different m/z peaks is indicated.

doi:10.1371/journal.pone.0135448.g002

2.58 Å) with Asp90 from the opposing dimer, which presumably stabilizes the tetramer arrangement. The other N-terminal peptide inserts into a cleft between neighboring tetramers in the crystal lattice. The dimers in the asymmetric unit interact substantially (1370 Å² accessible surface area buried on each dimer). Indeed, CrRbcX-IIa(34–156) formed mainly tetramers in solution as detected by native-MS (Fig 2B). However, this interaction is unlikely to be functionally relevant since full-length CrRbcX-IIa behaved as a dimer in solution (Fig 2A).

Comparison with Other RbcX Structures

The crystal structure of the dimer of CrRbcX-IIa(34–156) is closely similar to that of the plant ortholog AtRbcX-II (AtRbcX1) [18] (Fig 5A). 175 Cα positions could be superposed with a r.m.s.d. of 1.239 Å. In contrast, the structure of CrRbcX-IIa(34–156) differs more substantially from the structures of cyanobacterial RbcX and AtRbcX-I. For example, while one four-helix bundle and the associated C-terminal helix from the other subunit of the dimer of AtRbcX-I are reasonably well superimposable with CrRbcX-IIa(34–156) (r.m.s.d. 1.414 Å for 120 matching Cα atom positions), the other helical bundle is markedly shifted (Fig 5B). The situation is closely similar when comparing with the cyanobacterial *Anabaena* sp. CA RbcX (Anaca-RbcX) with an r.m.s.d. 1.453 Å for 134 matching Cα atom positions (Fig 5C). The rearrangement displaces helices α1 and α1' in the protomers longitudinally, which moves the symmetry-related pairs of hydrophobic pockets apart by ~5 Å. This becomes apparent from comparing the positions of residues Leu57 and Phe62, which line the hydrophobic pockets (spheres in Fig 5). Consequently, a pseudo-symmetrical binding of the FEF motif in the RbcL C-terminal peptide across the dyad axis is not possible in CrRbcX, in contrast to the binding mode of the FEF motif to cyanobacterial RbcX [10, 12]. The helices α2 of CrRbcX-IIa(34–156),

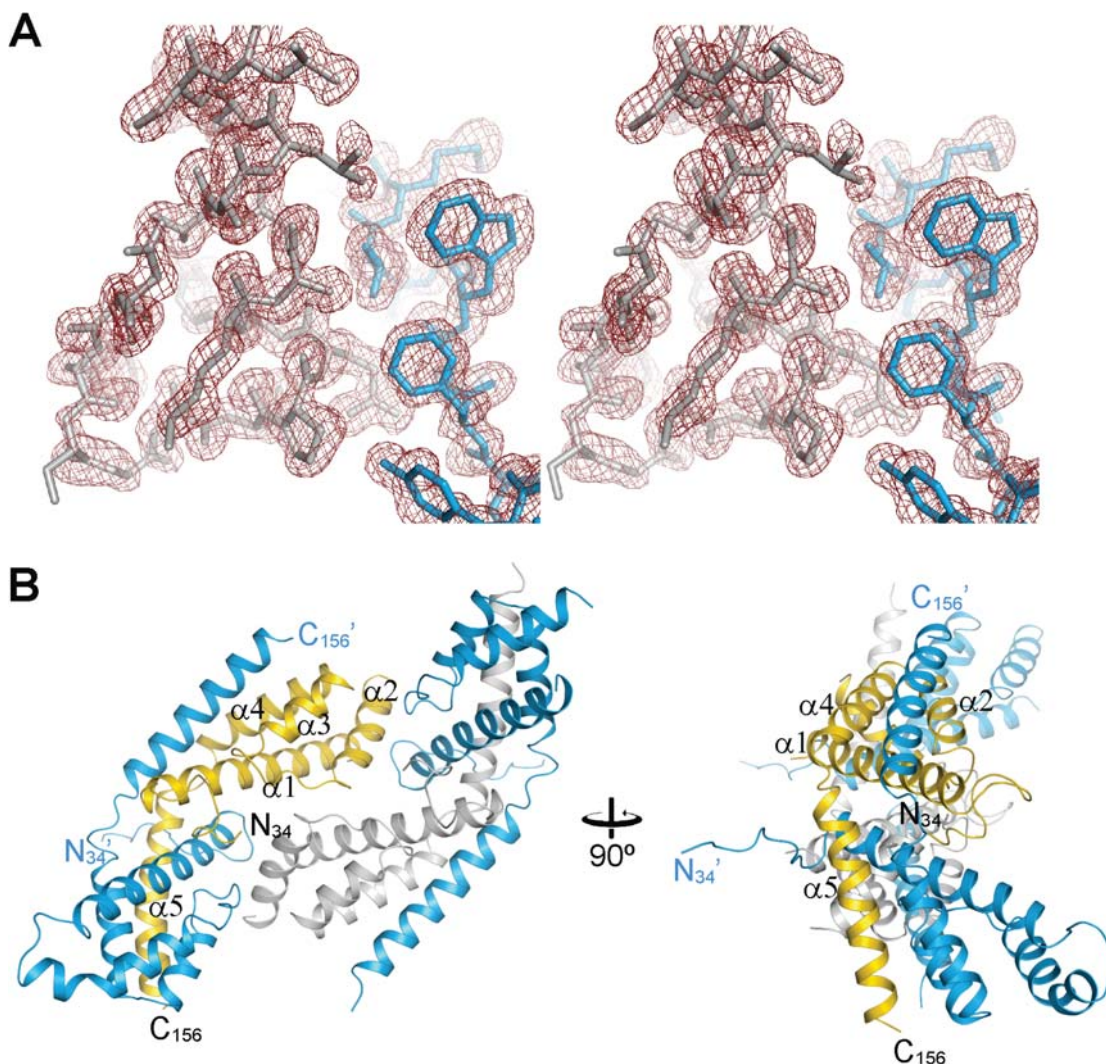


Fig 3. Asymmetric unit of the CrRbcX-IIa(34–156) crystal. (A) Stereoview of a representative portion of the experimental density map at 1.0 σ . The final model is superposed in stick representation. (B) Tetrameric complex of the SeMet-labeled oligomer in the asymmetric unit of the crystal lattice. Two perpendicular views are shown. On the left, a view along the two-fold molecular axis is shown. CrRbcX-IIa(34–156) is shown in ribbon representation. In each dimer, one of the chains is colored blue and the other gold or silver, respectively. Chain termini and secondary structure elements are indicated. The N-termini of the golden/silver subunits reach into the clefts (roughly horizontal); the N-termini of the blue subunits towards crystallographic symmetry mates.

doi:10.1371/journal.pone.0135448.g003

which form the “walls” of the hydrophobic cleft, are rotated outwards in comparison to cyanobacterial RbcX (Fig 5C), widening the cleft.

Structural Basis of RbcL Peptide Recognition

Attempts to obtain a co-crystal between CrRbcXIIa(34–159) and the C-terminal RbcL peptide failed, presumably due to low peptide binding affinity. Taking advantage of the finding that the N-terminus of RbcX binds into the central cleft (Fig 4), we therefore generated a fusion construct between CrRbcX-IIa and the C-terminal recognition motif in CrRbcL. In this construct, residues 462–473 of CrRbcL (sequence WKEIKFEFDTID) are directly linked to residue Pro37 at the N-terminus of CrRbcX-IIa(37–156), with the new N-terminus of the fusion protein starting with Trp462 of the RbcL sequence. This fusion protein readily crystallized and the structure

Table 1. Crystallographic data collection and model refinement statistics.

| Dataset | CrRbcX-IIa(34–156) (SeMet) | CrRbcL(462–474)-CrRbcX-IIa(37–156) |
|---|----------------------------|------------------------------------|
| Data collection | | |
| Wavelength (Å) | 0.9790 | 0.9999 |
| Space group | <i>P</i> 1 | <i>P</i> 1 |
| Cell dimensions | | |
| a, b, c (Å); | 36.13, 52.99, 61.56; | 34.22, 38.53, 50.36; |
| α, β, γ (°) | 76.49, 81.10, 70.10 | 88.47, 81.53, 67.92 |
| Resolution limits (Å)* | 59.66–1.6 (1.69–1.6) | 35.68–1.97 (2.07–1.97) |
| <i>R</i> _{merge} * | 0.059 (0.332) | 0.081 (0.494) |
| <i>I</i> /σ <i>I</i> * | 18.3 (4.6) | 10.4 (2.0) |
| Multiplicity * | 7.1 (6.7) | 2.4 (2.2) |
| Completeness (%) * | 94.1 (77.2) | 94.2 (87.9) |
| Wilson B-factor (Å ²) | 15.3 | 22.8 |
| Refinement | | |
| Resolution range | 30–1.6 | 30–1.97 |
| Reflections ** | 49112 (2585) | 15009 (1887) |
| <i>R</i> _{work} / <i>R</i> _{free} | 0.177 / 0.206 | 0.200 / 0.222 |
| Number of atoms | 4185 | 1887 |
| Average B-factor (Å ²) | 19.0 | 31.0 |
| r.m.s. deviations | | |
| Bond length (Å) | 0.011 | 0.013 |
| Bond angle (°) | 1.509 | 1.423 |
| Ramachandran plot *** | | |
| Favoured (%) | 97.7 | 97.3 |
| Allowed (%) | 2.1 | 2.2 |
| Outliers (%) | 0.2 | 0.4 |

* Values in parenthesis for outer shell.

** Values in parenthesis for test set.

*** Values from Molprobit 4.02.

doi:10.1371/journal.pone.0135448.t001

was solved at 1.97 Å resolution (Table 1). The structural core of CrRbcX-IIa(37–156) in the fusion protein is virtually identical to that obtained for CrRbcX-IIa(34–156) (r.m.s.d. 0.425 Å for 211 matching Cα positions). Thus it is unlikely that the contact area with the RbcL peptide is distorted by crystal packing. Difference electron density along the hydrophobic cleft could be assigned to the RbcL residues 462–467 (WKEIKF). Residues 468–473 (EFDTID) of RbcL as well as residues 37–43 of CrRbcX-IIa were disordered (Fig 6A). Notably, Phe469 was among the disordered residues, consistent with the finding that the corresponding Phe464 in Syn7002-RbcL is functionally less important for RbcX binding than Phe462 (Phe467 in CrRbcL) [12]. The sidechains of Ile465 and Phe467 point into hydrophobic pockets surrounded by Phe60/Arg64/Leu67/Leu92 and Leu57/Phe60/Met96, respectively (Fig 6B). The sidechain of Lys463 points towards the C-terminal end of helix α2 and Asp90. The indole moiety of Trp462 interacts with Tyr85 and Met89, but also with a neighboring CrRbcX-IIa molecule (not shown), and thus these interactions seem to be influenced by crystal packing.

Superposition with the structure of the heterologous cyanobacterial Syn6301-RbcL₈/Anaca-RbcX₈ assembly intermediate [10] shows that Ile465 and Phe467 of CrRbcL are recognized by similar sites on CrRbcX-IIa (Fig 6C). The peptide is oriented more towards helix α2 in the cyanobacterial structure, whereas it assumes a deeper and more central position in the

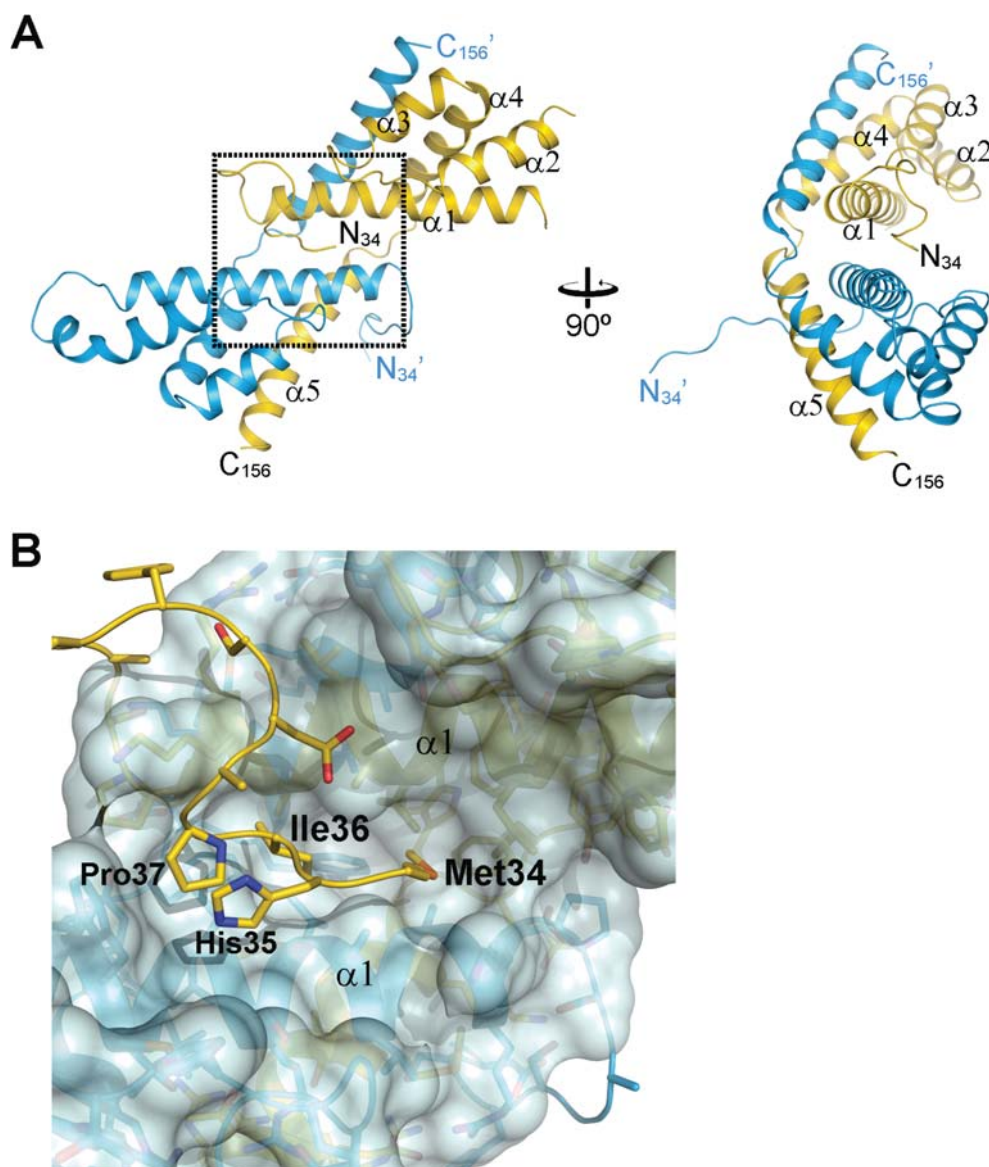


Fig 4. Crystal structure of the CrRbcX-IIa(34–156) dimer. (A) Ribbon representation of the CrRbcX-IIa(34–156) dimer. Two perpendicular views are shown, the first along the molecular two-fold axis. (B) Interactions of the N-terminal tail with the hydrophobic cleft in CrRbcX-IIa(34–156). A zoom-in on the boxed area in panel (A) is shown. The N-terminal tail is shown as a coil with prominent sidechains in stick representation. The bulk of the CrRbcX-IIa(34–156) is represented as a molecular surface.

doi:10.1371/journal.pone.0135448.g004

hydrophobic cleft of CrRbcX-IIa (Fig 6C). The indole ring of Trp462 is at roughly the same place in the superposition, but the backbone conformations differ strongly at this segment. We note that in the context of the RbcL subunit this residue would be connected, whereas it forms the N-terminal residue in the fusion construct. This difference in sequence topology may influence the binding mode.

The superposed CrRbcX-IIa is compatible with the surface of the RbcL anti-parallel dimer in the context of the RbcL₈ core complex (Fig 6D), in a topology similar to that observed for the cyanobacterial RbcX [10]. The C-terminal sequence of one RbcL subunit reaches into the central cleft of CrRbcX-IIa and the functionally critical, conserved residues Gln69 and Arg118

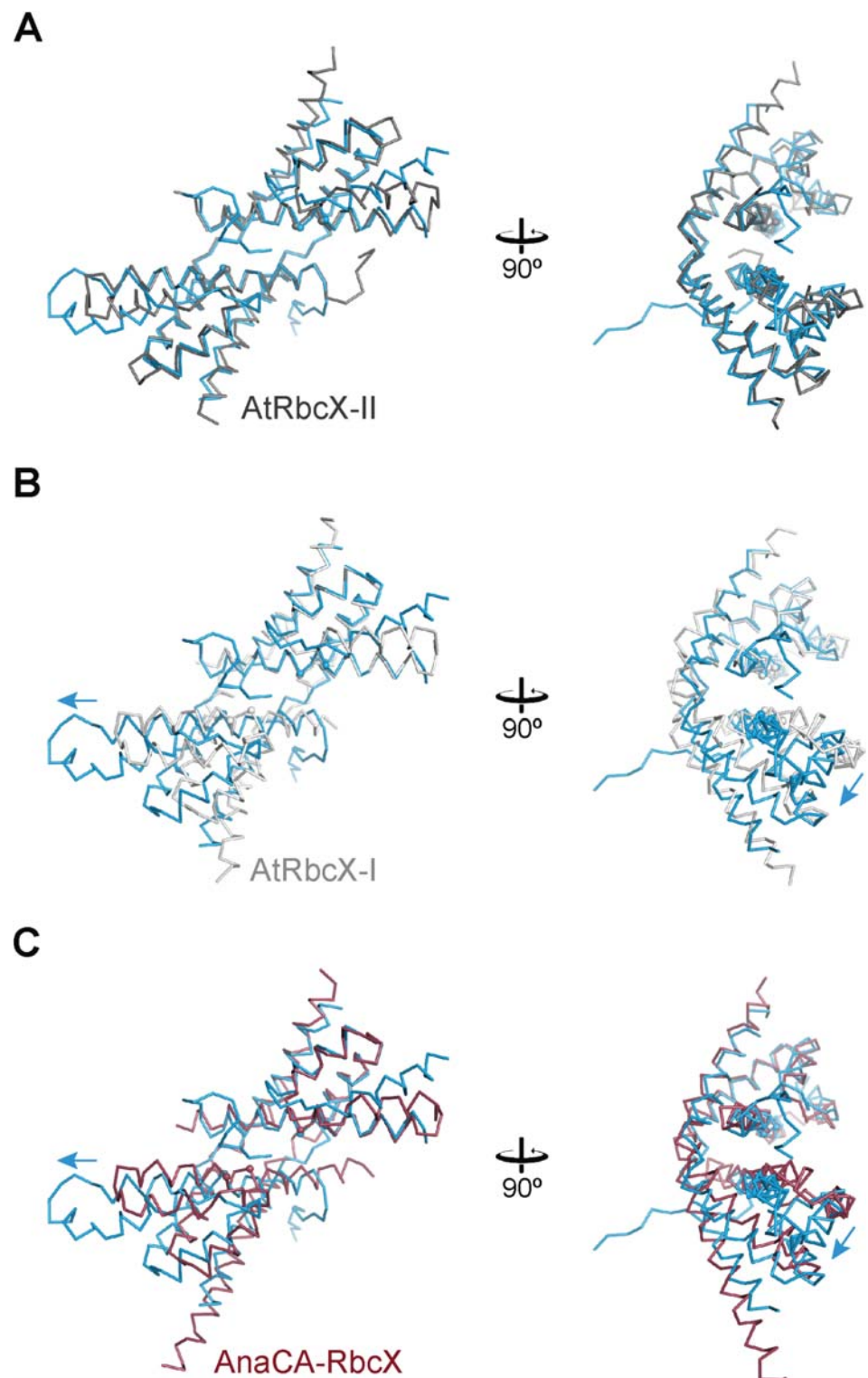


Fig 5. Comparison of the CrRbcX-IIa(34–156) structure with RbcX-II and RbcX-I homologs. (A) Comparison with the *A. thaliana* homolog AtRbcX-II. The backbones of the subunits of CrRbcX-IIa(34–156) are represented as a Ca trace in the same views as in Fig 4A. Spheres designate the Cα atoms of Leu57 and Phe60 in CrRbcX-IIa, or the respective sequence positions in the homologous proteins. CrRbcX-IIa and

AtRbcX-II are shown in blue and dark grey, respectively. (B) Comparison with the *A. thaliana* homolog AtRbcX-I, which is shown in light grey. (C) Comparison with cyanobacterial AnaCA-RbcX which is shown in red. Arrows indicate the direction of displacement of the second 4-helix bundle of CrRbcX-IIa(34–156).

doi:10.1371/journal.pone.0135448.g005

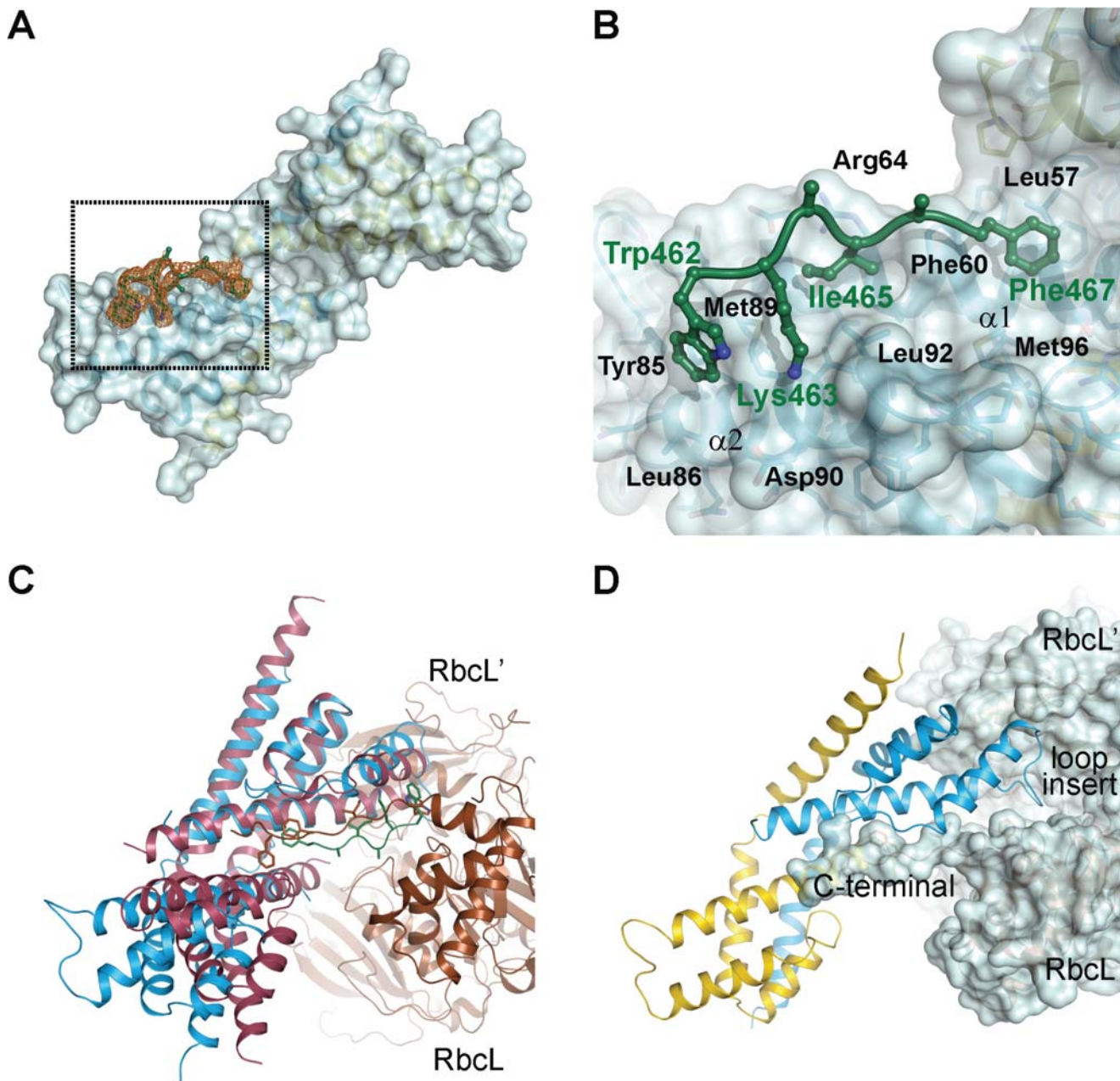


Fig 6. Crystal structure of a fusion protein revealing the interactions between CrRbcX-IIa and the C-terminal tail of CrRbcL. (A) Unbiased omit difference electron density for the RbcL tail residues of the CrRbcL(462–474)-RbcX-IIa(37–156) fusion protein. The C-terminal sequence of CrRbcL is shown as a coil and the sidechains in stick representation. The difference electron density at 1.5 σ level is shown as orange meshwork. CrRbcX-IIa(37–156) is represented as a molecular surface. (B) Detailed view of the RbcL-RbcX interactions. The area boxed in panel (A) is shown. (C) Superposition of the CrRbcX-IIa(37–156) onto the Syn6301-RbcL₈/AnaCA-RbcX₈ crystal structure [10]. The structures are shown in ribbon representation. The RbcL subunits are shown in brown and sienna; the AnaCA-RbcX dimer in red; CrRbcX-IIa dimer in blue. (D) Putative contacts of CrRbcX-IIa(37–156) with the surface of the Syn6301-RbcL₈ complex. The same view as in panel (C) is shown.

doi:10.1371/journal.pone.0135448.g006

(Fig 1) are positioned correctly for interaction with the second RbcL subunit (Fig 6D). The loop insertion between helices $\alpha 1$ and $\alpha 2$ of CrRbcX-IIa, which is ordered in the structure of the CrRbcL(462–474)-RbcX-IIa(37–156) fusion protein, would extend into a shallow groove of the RbcL dimer surface (Fig 6D). We speculate that this loop insertion found in RbcX sequences of green algae might modulate the interaction with RbcL.

Functional Characterization of CrRbcX

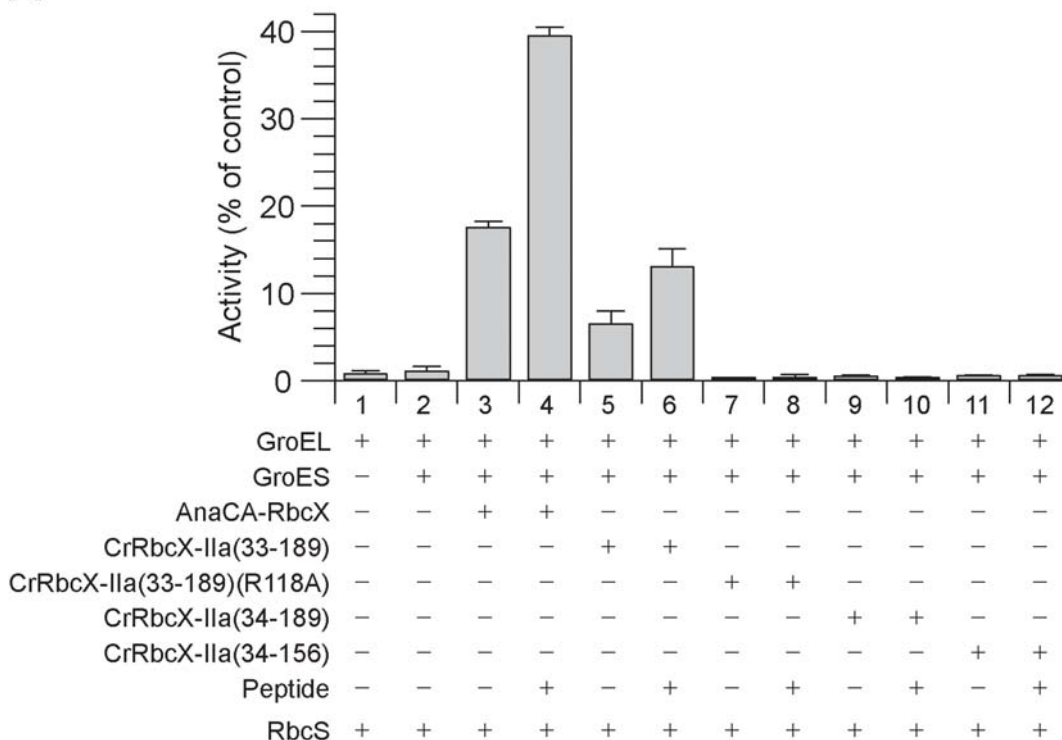
We used the previously reconstituted Rubisco from *S. elongatus* PCC6301 [9] and the bacterial chaperonin system GroEL/ES to assess the functionality of CrRbcX-IIa in Rubisco assembly. Unfolded RbcL was bound to GroEL upon dilution from denaturant. Assembly was initiated by adding GroES, ATP and RbcX for 60 min at 25°C, followed by addition of RbcS for Rubisco activity assay. The formation of holoenzyme was dependent on RbcX as shown previously [9], reaching a yield of ~20% with AnaCA-RbcX (Fig 7A). Addition of the C-terminal RbcL peptide prior to RbcS doubled the yield to ~40% by facilitating the displacement of RbcX from the RbcL₈RbcX₈ assembly intermediate by RbcS [9, 10]. A lower yield of enzyme activity of ~7% was obtained with full-length CrRbcX-IIa(33–189) protein, but only when present at a high molar excess (30 μ M dimer) over RbcL. Again the activity doubled in the presence of the C-terminal RbcL peptide (Fig 7A). The mutant CrRbcX-IIa(R118A) did not support assembly, consistent with this conserved residue being involved in the stabilization of the RbcL dimer [9, 10, 12]. Notably, CrRbcX-IIa(34–189), lacking the N-terminal residue Arg33 of the full-length protein was inactive (Fig 7A). Arg or Lys is conserved at this position among most RbcX-II homologs (Fig 1). Arg33 is also missing in the C-terminally truncated, crystallized CrRbcX-IIa(34–156) protein. In the crystal structure, the amino group of the N-terminal Met34 forms a short salt bridge (2.53–2.58 Å) with Asp90 from the other RbcX dimer, which appears to stabilize the tetramer. In addition, Arg33 would clash with the other dimer, consistent with the MS data showing that deletion of Arg33 favors tetramer formation in solution (Fig 2). We suggest that in the absence of Arg33, the N-terminus of RbcX may bind into the central cleft, rendering the protein non-functional in Rubisco assembly (Fig 7A), consistent with the formation of non-functional tetramers (Fig 7B).

Discussion

Our data demonstrate that RbcX-II from the green algae *C. reinhardtii* functions as a bona fide Rubisco assembly chaperone, despite its considerable evolutionary distance from cyanobacterial and eukaryotic RbcX-I proteins. Like all other known RbcX proteins, CrRbcX-IIa is an arc-shaped dimer with a central hydrophobic cleft that binds the C-terminal sequence of the RbcL subunit. Conserved polar residues at the corners of RbcX make critical contacts to the N-domain of an adjacent RbcL, thereby stabilizing the RbcL anti-parallel dimer in a state competent for assembly to the RbcL₈ core complex of Rubisco.

The crystal structure of CrRbcX-IIa differs from the structures of cyanobacterial RbcX homologs in several aspects. The adjacent helices $\alpha 1$, which form the floor of the hydrophobic cleft, are shifted with respect to each other, moving the binding pockets for the two Phe side-chains in the C-terminal RbcL binding motif apart. Consistently, density for the bound peptide sequence is only discernible until the first Phe residue (Phe467) in the complex structure. There are additional hydrophobic cavities between the helices close to the symmetry axis, resulting from the conserved substitution of Thr10 in cyanobacterial RbcX by Ala in RbcX-II (residue 50 in CrRbcX-IIa sequence numbering), but these volumes are not occupied in the complex with peptide. In the apo-structure, the sidechains of the conserved residues Met34 and Ile36 point into these pockets, but the functional significance of this interaction, if any, is

A



B

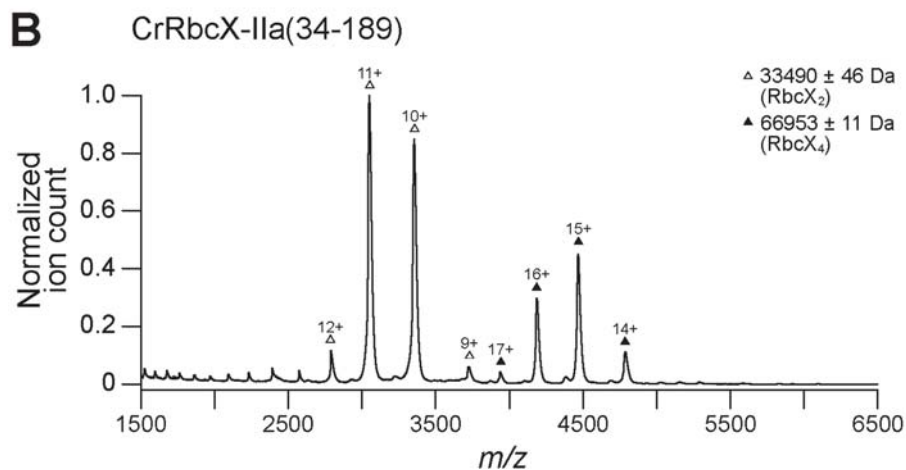


Fig 7. Rubisco reconstitution of CrRbcX-IIa and oligomeric state of CrRbcX-IIa(34–189) analyzed by native-MS. (A) Rubisco reconstitution. Chemically denatured RbcL from *S. elongatus* PCC6301 (at 100 μ M) was diluted 200-fold into ice-cold buffer containing GroEL (1.0 μ M). The components (2 μ M GroES oligomer; 2 μ M AnaCa-RbcX or 30 μ M CrRbcX dimer) were added as indicated and refolding/assembly initiated by addition of 4 mM ATP at 25°C (see [Materials and Methods](#)). After incubation for 60 min, RbcS (5 μ M) was added with or without C-terminal RbcL peptide (200 μ M) for 15 min, followed by Rubisco enzyme assay. The activity of RbcL₈ core complex (\sim 0.06 μ M oligomer) incubated with RbcS (5 μ M) was set to 100%. Error bars s.d. (n = 3 independent experiments). (B) Nano-ESI native-MS spectra of CrRbcX-IIa(34–189). Symbols indicate the charge state distributions with the charge states shown for some peaks; the calculated mass around the m/z values of the respective protein complexes is reported. The accuracy of mass values calculated from the different m/z peaks is indicated.

doi:10.1371/journal.pone.0135448.g007

unclear. Interestingly, in the structure of the *A. thaliana* ortholog, which has essentially the same backbone conformation, the pockets are smaller and intra-molecular binding of the N-terminus into the central cleft is not observed.

Besides the RbcX homologs, a recent screen of a Maize mutant library identified several additional Rubisco accumulation factors, including Bsd2, Raf1 and Raf2 [38–42]. RbcX and Raf1 are generally conserved in photosynthetic organisms containing form IB Rubisco [2, 3], but mediate assembly by different mechanisms [43]. Whether RbcX and Raf1 cooperate in a coherent assembly pathway or act in parallel pathways is still unknown.

Acknowledgments

C. reinhardtii cDNA was obtained from the Kazusa Institute and from E. Lorentzen. We thank R. Körner for performing the mass spectroscopy analysis of the protease-treated CrRbcX-IIa, J. Y. Bhat and D. Balchin for the native-MS analysis. We also thank the staff at the Core and Crystallization Facilities at the MPI of Biochemistry for their services and acknowledge support by the staff at beamline X10SA of the Swiss Light Source (SLS) in Villigen, Switzerland.

Author Contributions

Conceived and designed the experiments: AB MH-H. Performed the experiments: AB TH CL. Analyzed the data: AB TH CL FUH MH-H. Contributed reagents/materials/analysis tools: AB TH CL. Wrote the paper: AB FUH MH-H.

References

1. Andersson I, Backlund A. Structure and function of Rubisco. *Plant Physiol Biochem.* 2008; 46(3):275–91. PMID: [18294858](#). doi: [10.1016/j.plaphy.2008.01.001](#)
2. Tabita FR. Microbial Ribulose 1,5-bisphosphate carboxylase/oxygenase: a different perspective. *Photosynth Res.* 1999; 60(1):1–28. doi: [10.1023/a:1006211417981](#)
3. Tabita FR, Satagopan S, Hanson TE, Kreef NE, Scott SS. Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *J Exp Bot.* 2008; 59(7):1515–24. PMID: [182817172](#). doi: [10.1093/jxb/ern361](#)
4. Andersson I. Catalysis and regulation in Rubisco. *J Exp Bot.* 2008; 59(7):1555–68. PMID: [18417482](#). doi: [10.1093/jxb/ern091](#)
5. Hauser T, Popilka L, Hartl FU, Hayer-Hartl M. Role of auxiliary proteins in Rubisco biogenesis and function. *Nature Plants.* 2015; 1. doi: [10.1038/nplants.2015.65](#)
6. Strittmatter P, Soll J, Bölder B. The chloroplast protein import machinery: a review. *Methods Mol Biol.* 2010; 619:307–21. PMID: [20419418](#). doi: [10.1007/978-1-60327-412-8_18](#)
7. Barraclough R, Ellis RJ. Protein synthesis in chloroplasts. IX. Assembly of newly-synthesized large subunits into ribulose bisphosphate carboxylase in isolated intact pea chloroplasts. *Biochim Biophys Acta.* 1980; 608(1):18–31. PMID: [7388030](#).
8. Hemmingsen SM, Woolford C, van der Vies SM, Tilly K, Dennis DT, Georgopoulos CP, et al. Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature.* 1988; 333(6171):330–4. PMID: [2897629](#).
9. Liu C, Young AL, Starling-Windhof A, Bracher A, Saschenbrecker S, Rao BV, et al. Coupled chaperone action in folding and assembly of hexadecameric Rubisco. *Nature.* 2010; 463(7278):197–202. PMID: [20075914](#). doi: [10.1038/nature08651](#)
10. Bracher A, Starling-Windhof A, Hartl FU, Hayer-Hartl M. Crystal structure of a chaperone-bound assembly intermediate of form I Rubisco. *Nat Struct Mol Biol.* 2011; 18(8):875–80. PMID: [21765418](#). doi: [10.1038/nsmb.2090](#)
11. Onizuka T, Endo S, Akiyama H, Kanai S, Hirano M, Yokota A, et al. The *rbcX* gene product promotes the production and assembly of Ribulose-1,5-bisphosphate carboxylase/oxygenase of *Synechococcus* sp. PCC7002 in *Escherichia coli*. *Plant Cell Physiol.* 2004; 45(10):1390–5. PMID: [15564522](#).
12. Saschenbrecker S, Bracher A, Rao KV, Rao BV, Hartl FU, Hayer-Hartl M. Structure and function of RbcX, an assembly chaperone for hexadecameric Rubisco. *Cell.* 2007; 129(6):1189–200. PMID: [17574029](#).

13. Larimer FW, Soper TS. Overproduction of Anabaena 7120 Ribulose-bisphosphate carboxylase/oxygenase in *Escherichia coli*. *Gene*. 1993; 126(1):85–92. PMID: [8472962](#).
14. Emlyn-Jones D, Woodger FJ, Price GD, Whitney SM. RbcX can function as a rubisco chaperonin, but is non-essential in *Synechococcus* PCC7942. *Plant Cell Physiol*. 2006; 47(12):1630–40. PMID: [17071623](#).
15. Tabita FR. Rubisco: The enzyme that keeps on giving. *Cell*. 2007; 129(6):1039–40. PMID: [17574015](#).
16. Tanaka S, Sawaya MR, Kerfeld CA, Yeates TO. Structure of the Rubisco chaperone RbcX from *Synechocystis* sp. PCC6803. *Acta Crystallogr D Biol Crystallogr*. 2007; 63(Pt 10):1109–12. PMID: [17881829](#).
17. Tarnawski M, Krzywda S, Bialek W, Jaskolski M, Szczepaniak A. Structure of the RuBisCO chaperone RbcX from the thermophilic cyanobacterium *Thermosynechococcus elongatus*. *Acta Crystallogr F Struct Biol Cryst Commun*. 2011; 67(Pt 8):851–7. PMID: [21821880](#).
18. Kolesinski P, Golik P, Grudnik P, Piechota J, Markiewicz M, Tarnawski M, et al. Insights into eukaryotic Rubisco assembly—crystal structures of RbcX chaperones from *Arabidopsis thaliana*. *Biochim Biophys Acta*. 2013; 1830(4):2899–906. PMID: [23295968](#). doi: [10.1016/j.bbagen.2012.12.025](#)
19. Kolesinski P, Piechota J, Szczepaniak A. Initial characteristics of RbcX proteins from *Arabidopsis thaliana*. *Plant Mol Biol*. 2011; 77(4–5):447–59. PMID: [21922322](#). doi: [10.1007/s11103-011-9823-8](#)
20. Asamizu E, Nakamura Y, Miura K, Fukuzawa H, Fujiwara S, Hirono M, et al. Establishment of publicly available cDNA material and information resource of *Chlamydomonas reinhardtii* (Chlorophyta) to facilitate gene function analysis. *Phycologia*. 2004; 43(6):722–6. doi: [10.2216/i0031-8884-43-6-722.1](#)
21. Baker RT, Catanzariti AM, Karunasekara Y, Soboleva TA, Sharwood R, Whitney S, et al. Using deubiquitylating enzymes as research tools. *Methods Enzymol*. 2005; 398:540–54. PMID: [16275357](#).
22. Brinker A, Pfeifer G, Kerner MJ, Naylor DJ, Hartl FU, Hayer-Hartl M. Dual function of protein confinement in chaperonin-assisted protein folding. *Cell*. 2001; 107(2):223–33. PMID: [11672529](#).
23. Van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J Mol Biol*. 1993; 229(1):105–24. PMID: [7678431](#).
24. Goloubinoff P, Gatenby AA, Lorimer GH. GroE heat-shock proteins promote assembly of foreign prokaryotic Ribulose bisphosphate carboxylase oligomers in *Escherichia coli*. *Nature*. 1989; 337(6202):44–7. PMID: [2562907](#).
25. Kabsch W. XDS. *Acta Crystallogr D Biol Crystallogr*. 2010; 66(Pt 2):125–32. PMID: [20124692](#). doi: [10.1107/S0907444909047337](#)
26. Evans P. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr*. 2006; 62(Pt 1):72–82. PMID: [16369096](#).
27. Evans PR. Scala. Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography. 1997; 33:22–4.
28. French G, Wilson K. On the treatment of negative intensity observations. *Acta Crystallogr Section A*. 1978; 34(4):517–25. doi: [10.1107/S0567739478001114](#)
29. Collaborative Computational Project N. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*. 1994; 50(Pt 5):760–3. PMID: [15299374](#).
30. Pape T, Schneider TR. HKL2MAP: a graphical user interface for phasing with SHELX programs. *J Appl Cryst*. 2004; 37(5):843–4. doi: [10.1107/S0021889804018047](#)
31. Sheldrick GM. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr*. 2010; 66(Pt 4):479–85. PMID: [20383001](#). doi: [10.1107/S0907444909038360](#)
32. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*. 2004; 60(Pt 12 Pt 1):2126–32. PMID: [15572765](#).
33. Murshudov GN, Vagin AA, Dodson EJ. Refinement of Macromolecular Structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*. 1997; 53(Pt 3):240–55. PMID: [15299926](#).
34. Vagin AA, Isupov MN. Spherically averaged phased translation function and its application to the search for molecules and fragments in electron-density maps. *Acta Crystallogr D Biol Crystallogr*. 2001; 57(Pt 10):1451–6. PMID: [11567159](#).
35. Kleywegt GT, Jones TA. A super position. Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography. 1994; 31:9–14.
36. Schrödinger LLC. The PyMOL Molecular Graphics System, Version 1.3. 2010.
37. Gouet P, Courcelle E, Stuart DI, Metoz F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*. 1999; 15(4):305–8. PMID: [10320398](#).

38. Feiz L, Williams-Carrier R, Wostrickoff K, Belcher S, Barkan A, Stern DB. Ribulose-1,5-bis-phosphate carboxylase/oxygenase accumulation factor1 is required for holoenzyme assembly in maize. *Plant Cell*. 2012; 24(8):3435–46. PMID: [22942379](#).
39. Wheatley NM, Sundberg CD, Gidaniyan SD, Cascio D, Yeates TO. Structure and identification of a pterin dehydratase-like protein as a ribulose-bisphosphate carboxylase/oxygenase (RuBisCO) assembly factor in the alpha-carboxysome. *J Biol Chem*. 2014; 289(11):7973–81. PMID: [24459150](#). doi: [10.1074/jbc.M113.531236](#)
40. Kolesinski P, Belusiak I, Czarnocki-Cieciura M, Szczepaniak A. Rubisco Accumulation Factor 1 from *Thermosynechococcus elongatus* participates in the final stages of ribulose-1,5-bisphosphate carboxylase/oxygenase assembly in *Escherichia coli* cells and in vitro. *FEBS J*. 2014; 281(17):3920–32. PMID: [25041569](#). doi: [10.1111/febs.12928](#)
41. Feiz L, Williams-Carrier R, Belcher S, Montano M, Barkan A, Stern DB. A protein with an inactive pterin-4a-carbinolamine dehydratase domain is required for Rubisco biogenesis in plants. *Plant J*. 2014; 80(5):862–9. PMID: [25279696](#). doi: [10.1111/tpj.12686](#)
42. Whitney SM, Birch R, Kelso C, Beck JL, Kapralov MV. Improving recombinant Rubisco biogenesis, plant photosynthesis and growth by coexpressing its ancillary RAF1 chaperone. *Proc Natl Acad Sci USA*. 2015; 112(11):3564–9. PMID: [25733857](#). doi: [10.1073/pnas.1420536112](#)
43. Hauser T, Bhat JY, Milićić G, Wendler P, Hartl FU, Bracher A, et al. Structure and mechanism of the Rubisco-assembly chaperone Raf1. *Nat Struct Mol Biol*. 2015; In Press.

3. Additional unpublished results

3.1 Analysis of Raf1-RbcL interaction by RbcL peptide membrane

The binding of Raf1 to the Rubisco large subunit was investigated by an array of techniques in Hauser *et al.*²⁰⁵. Most importantly, chemical cross-linking in combination with mass spectrometry provided detailed insight into the interaction of Raf1 with RbcL. To identify such sequence element(s) in RbcL, an array of 13-amino-acid-long, acetylated peptides covering the entire sequence of RbcL with an eleven residue overlap was C-terminally attached to a cellulose-PEG membrane. Incubation with Raf1, followed by detection with anti-Raf1 antibody, resulted in strong binding signals to specific peptides of the interaction partner. Such binding regions should normally show several spots next to each other since adjacent peptide sequences of RbcL overlap by 10 residues and binding should involve less than 11 residues. Therefore, cyanobacterial and plant Raf1 as well as their respective $\alpha+\beta$ domains were incubated with their homologous RbcL peptide sequence membranes, i.e. Syn6301-RbcL and At-RbcL, respectively (Fig.19, 21, 22 and 24). Membranes were washed after incubation several times, and binding was analyzed by western blotting against the respective Raf1 protein. As a control, the background signal of primary and secondary antibody binding to the membrane in absence of Raf1 was tested beforehand and no staining was observed for either anti 7942-Raf1/anti AtRaf1.2 or anti rabbit IgG (data not shown). Before each binding experiment, membranes were stripped and regenerated and background binding was tested to ensure that no signal from the previous incubation remained (data not shown). After incubation of the Syn6301-RbcL peptide membrane with Syn7942-Raf1, six discrete binding regions with at least 3 adjacent peptide spots were detected (Fig.19, left). Corresponding peptide motifs in overlapping binding regions are shown in red (Fig.19, right) and are highlighted in the Syn6301-RbcL antiparallel dimer (Fig.19). Interestingly, 3 peptide regions harbored Lysine residues that were found in Raf1-RbcL X-links before (Fig.20, purple)²⁰⁵. These peptides were located at the horizontal edges of the anti-parallel RbcL dimer, close to the region where Raf1 likely brackets the dimer. Thus, this result fits well with the proposed Raf1 mechanism in stabilizing RbcL antiparallel dimers. The other 3 peptide regions (Fig.20, sand), where no Raf1-RbcL X-links were found, all lie at the back of the dimer and are not in accordance with the model. These regions are in the dimer-dimer interaction side and would be masked when the RbcL₈ core is assembled. Raf1 stays bound to the RbcL₈ core and would not be able to bind to these regions simultaneously. Hence, these binding regions most likely do not play a role in Raf1 mediated Rubisco assembly. Since only short peptides are spotted

on the membrane it is very likely that additional false positive binding regions are detected and it is important to validate binding results with additional methods.

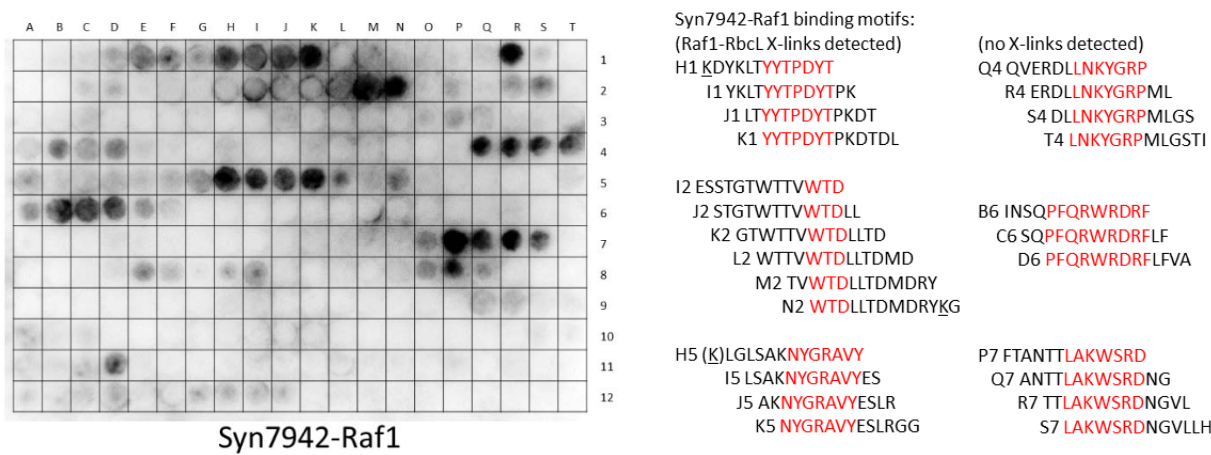


Fig.19: Syn7942-Raf1 binding to Syn6301-RbcL peptide membrane. A cellulose membrane containing an array of overlapping tridecamer peptides covering the sequence of Syn6301-RbcL was probed with Syn7942-Raf1. Peptide-bound Raf1 was visualized by chemiluminescent immunodetection with anti-Syn7942-Raf1 antibody. (left). Peptides where Syn7942-Raf1 was bound are indicated and overlapping binding sites of different peptides are highlighted in red (right). Lysine residues that were found in Raf1-RbcL X-links (Hauser *et al.*²⁰⁵) are underlined.

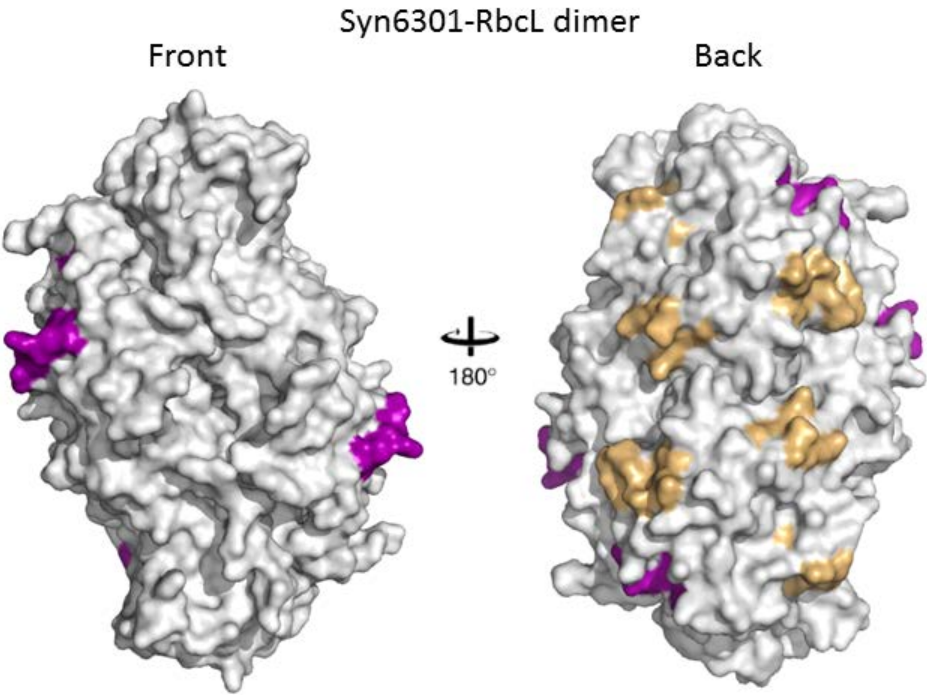


Fig.20: Syn7942-Raf1 binding to Syn6301-RbcL. Binding regions from Fig.19, left are mapped on the antiparallel Syn6301-RbcL dimer (white). Peptide regions that were found in Raf1-RbcL X-links are highlighted in purple, other peptides in sand. Left, front view of the Syn6301-RbcL antiparallel dimer, right, back view.

In contrast to full-length Syn7942-Raf1, both the respective α - and β -domain did show no significant binding to the Syn6301-RbcL peptide membrane, underlining the inability of the respective domains to interact with RbcL on their own (Fig.21)²⁰⁵.

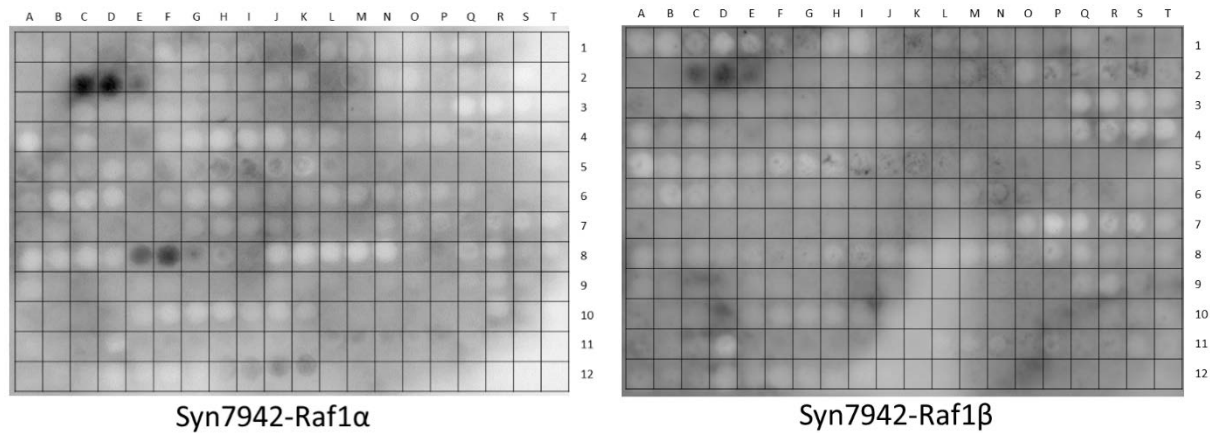


Fig.21: Syn7942-Raf1 domain binding to Syn6301-RbcL peptide membrane. Syn7942-Raf1 α (left) -and β (right) domains were incubated on a membrane consisting of spotted peptides of the Syn6301-RbcL sequence and analyzed by anti-Syn7942-Raf1 immunoblot.

To identify potential chaperonin binding sites, the Syn6301-RbcL peptide membrane was incubated with GroEL. Interestingly, GroEL showed distinct binding to a peptide motif on the Syn6301 Rubisco large sunbunit membrane (Fig.22, left). Moreover, the GroEL binding region lies in close proximity to one of the binding patches of Raf1 which was also found in cross-linking experiments which might explain their mutually exclusive binding ability (Fig.22, right).

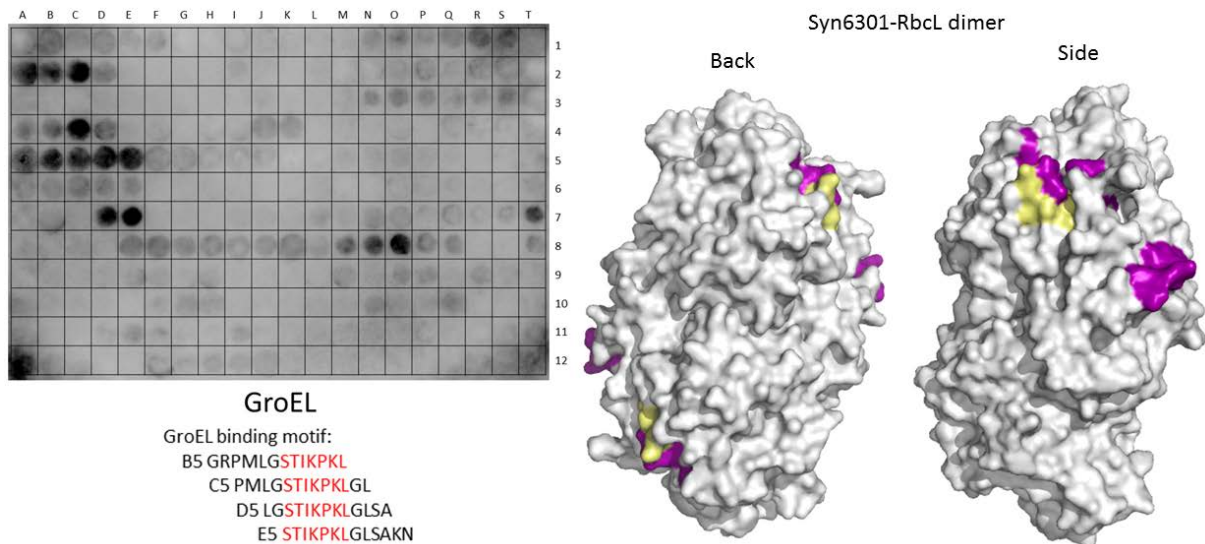


Fig.22: GroEL binding to Syn6301-RbcL peptide membrane. Left: GroEL was incubated on a membrane consisting of spotted peptides of the Syn6301-RbcL sequence and analyzed by anti-GroEL immunoblot. Peptides where GroEL was bound are indicated and overlapping binding sites of different peptides are highlighted in red. Right: GroEL binding regions are mapped on the antiparallel Syn6301-RbcL dimer (white) and are highlighted in pale yellow. Binding regions from Syn7942-Raf1 are shown in purple. Left, back view, right, side view. Note that binding region A2, B2 and C2 derive from anti GroEL antibody and are not specific to GroEL binding (data not shown).

When AtRaf1.2 was incubated on the At-RbcL peptide membrane eight discrete binding regions with at least 3 adjacent peptide spots were found, amongst them 3 strong regions where Raf1 signal intensity was considerably higher compared to the 6 weak binding regions (Fig.23, left). Corresponding binding motifs in overlapping binding regions are shown in red (Fig.23, right) and are highlighted in the *Spinacia oleracea*-RbcL antiparallel dimer in purple for the three identified strong binding regions (Fig.24).

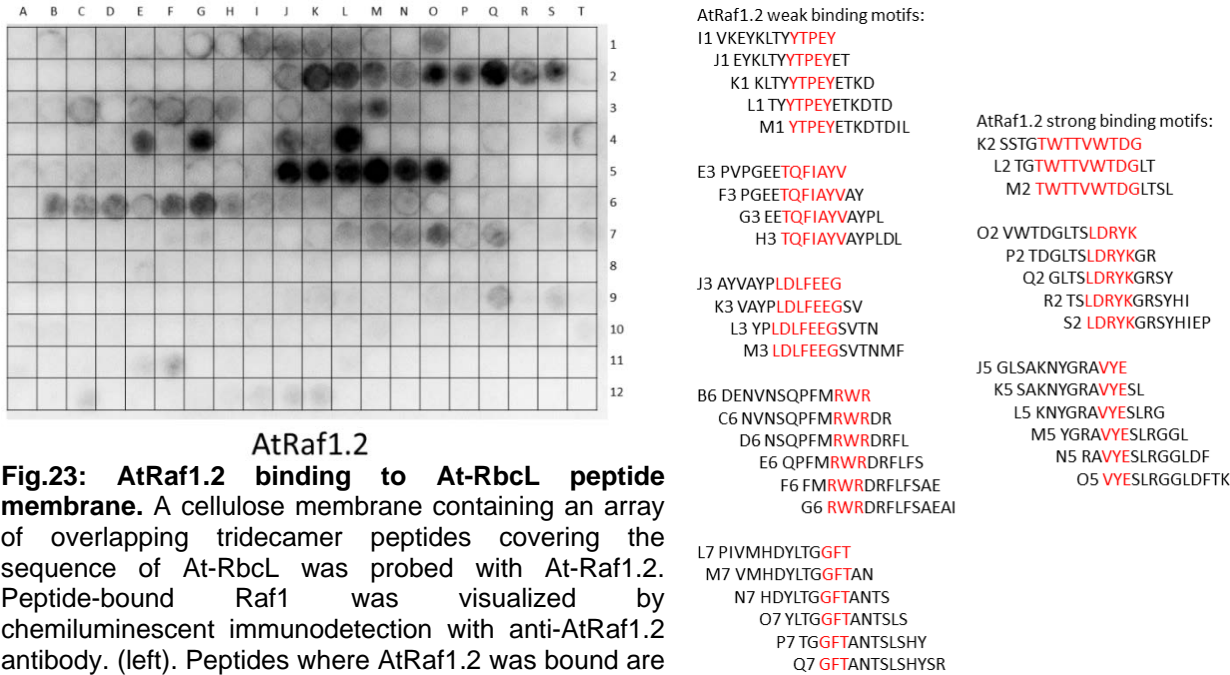


Fig.23: AtRaf1.2 binding to At-RbcL peptide membrane. A cellulose membrane containing an array of overlapping tridecamer peptides covering the sequence of At-RbcL was probed with At-Raf1.2. Peptide-bound Raf1 was visualized by chemiluminescent immunodetection with anti-AtRaf1.2 antibody. (left). Peptides where AtRaf1.2 was bound are indicated and overlapping binding sites of different peptides are highlighted in red (right).

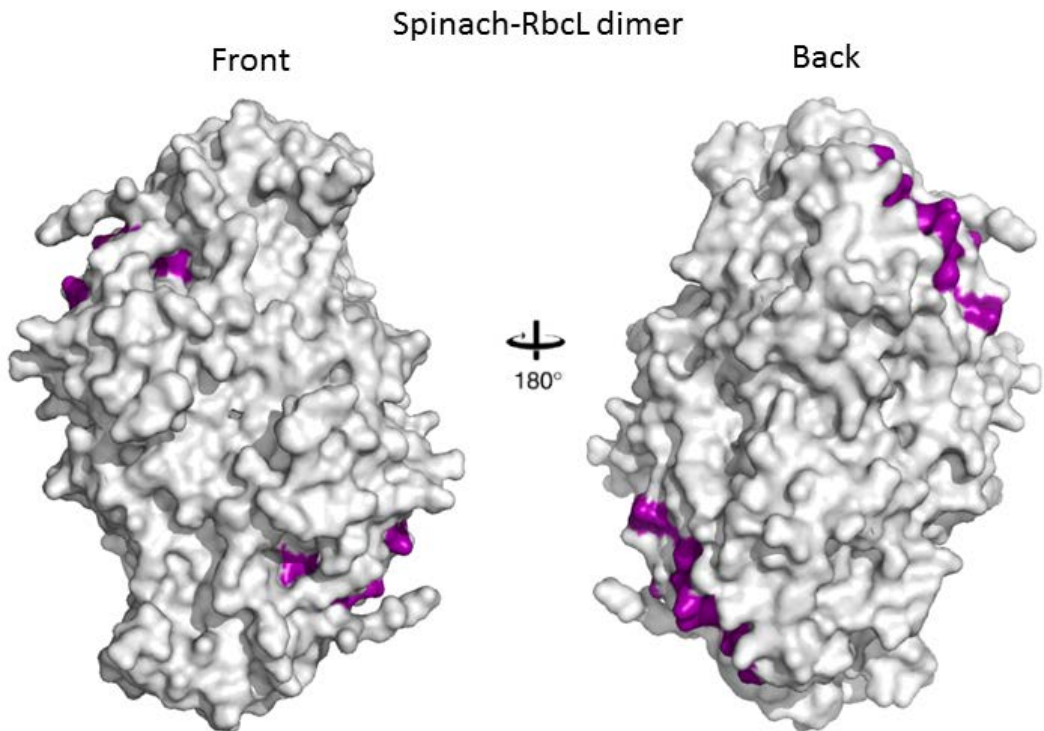


Fig.24: AtRaf1.2 binding to At-RbcL. Binding regions from Fig.22, left are mapped on the antiparallel Spinach-RbcL dimer (white) and highlighted in purple. Left, front view of the Spinach-RbcL antiparallel dimer, right, back view.

The three major binding regions of AtRaf1.2 on the At-RbcL peptide membrane are similar to the ones found for Syn7942-Raf1 on the Syn6301-RbcL peptide membrane. Binding occurs at the sides of the anti-parallel RbcL dimer which, is in accordance with the Raf1 mechanism in bracketing and stabilizing dimers. In contrast to full-length, similar to the Syn7942-Raf1 domains, both the AtRaf1.2 α - and β -domain did show no significant binding to the AtRaf1.2-RbcL peptide membrane, underlining the inability to interact with RbcL on their own (Fig.25)²⁰⁵. Taken together, binding of Raf1 to RbcL peptide membranes shows results consistent with other, more detailed and sensitive biophysical methods²⁰⁵. Therefore, it is important to validate binding to certain peptide regions by other techniques, as false positive results make convincing conclusions impossible.

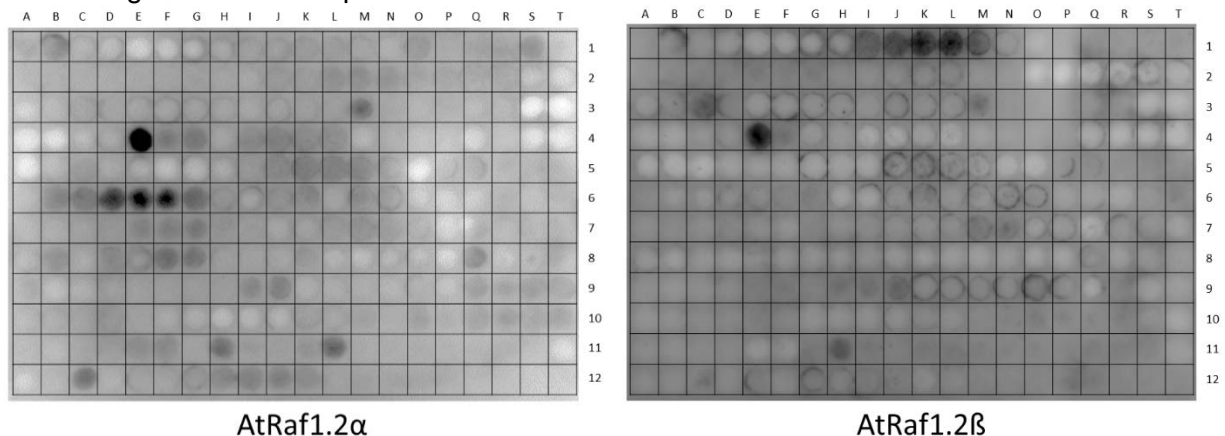


Fig.25: At-Raf1.2 domain binding to At-RbcL peptide membrane. At-Raf1.2 α (left) -and β (right) domains were incubated on a membrane consisting of spotted peptides of the At-RbcL sequence and analyzed by anti-AtRaf1.2 immunoblot.

3.2 Raf1-RbcL binding is not affected by Mg^{2+} , ATP and RuBP

Rubisco requires the binding of Mg^{2+} at its active site after carbamylation of Lysine 201 for its proper catalytic function. To test whether the binding of Mg^{2+} to purified RbcL₈ cores changes the stability of RbcL₈ in the presence and/or absence of GroEL/ATP and thus the binding of Raf1 to L₈ cores, RbcL₈ was incubated as before with different components (Fig.26)²⁰⁵. Overall, the combination of Mg^{2+} /ATP alone or in combination with GroEL did not change the stability of the RbcL₈ core complex (Fig.26, lanes 3 and 5). Furthermore, Raf1 binding to RbcL₈ was not altered in these conditions and two distinct species (RbcL* and a HMW complex) were observed as before. Thus, the RbcL₈ core is stable under Mg^{2+} /ATP conditions and does not bind to GroEL. As described before, RuBP binding to the Mg^{2+} -bound active site of Rubisco results in the closure of a series of loops (most importantly loop 6) over the site to enfold and capture the bisphosphate substrate. Closure of the loops brings together amino acids that are critical for catalysis and determine the fate of the substrate. Furthermore the very C-terminus of Rubisco closes the active site and renders it catalytically competent. Substrate binding has been reported to stabilize the RbcL antiparallel dimer and there are two binding sites formed at the subunit interface^{143,206,207,208}.

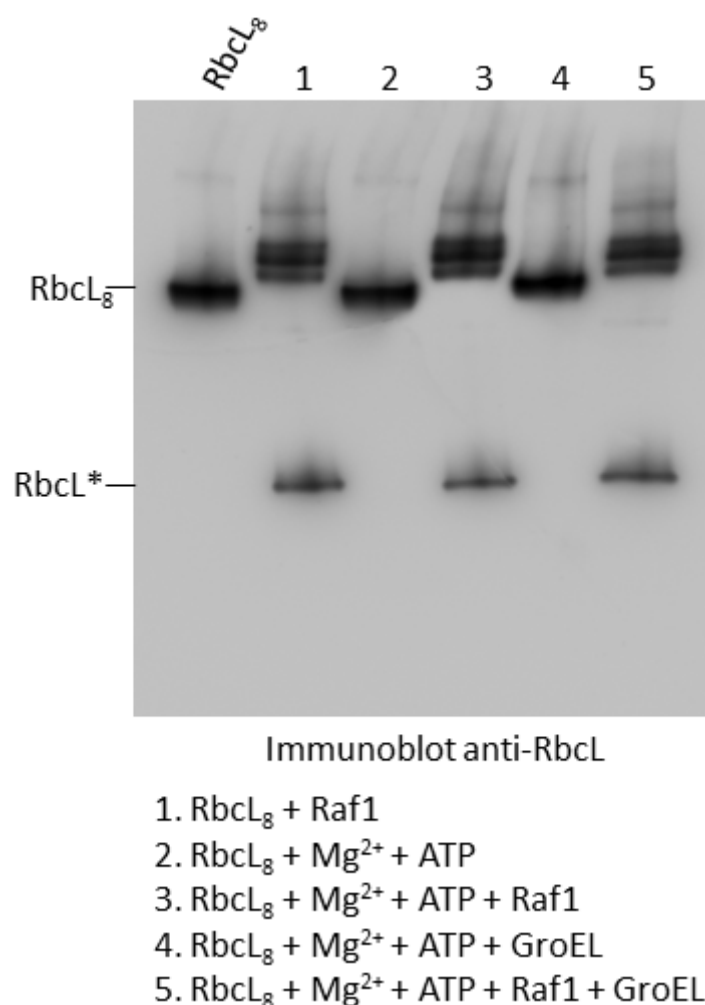


Fig.26: Interaction of Syn7942-Raf1 with pre-formed RbcL₈ complexes. Complex formation of RbcL₈ with Syn7942-Raf1, analyzed by native PAGE and immunoblotting with anti-RbcL. Samples are purified RbcL₈, incubated with Raf1, Mg²⁺, ATP or GroEL as indicated.

Therefore, the potential effect of additional RuBP Rubisco stabilization was tested in the *in vitro* reconstitution reaction as well as in the binding to RbcL₈ (Fig.27). In the presence of RuBP and RbcS (Fig.27, lane 1), no additional complex formation, e.g RbcL dimers or RbcL₈S₈ were observed as compared to the reaction w/o RuBP (Fig.27, lane 2). Moreover, no further stabilization of assembly complexes was observed in the presence of Raf1 and presence or absence of RbcS (Fig.27, lanes 2-6). Likewise, RuBP did not influence Raf1 binding to RbcL₈ and subsequent complex formation (Fig.27, lanes 7,8). Thus, RuBP does not have a chemical chaperone effect on cyanobacterial Rubisco dimer formation as observed for red-type bacterial form ID Rubisco²⁰⁶. Raf1 is essential and sufficient for binding and stabilization of folded RbcL against re-binding to GroEL as observed before²⁰⁵.

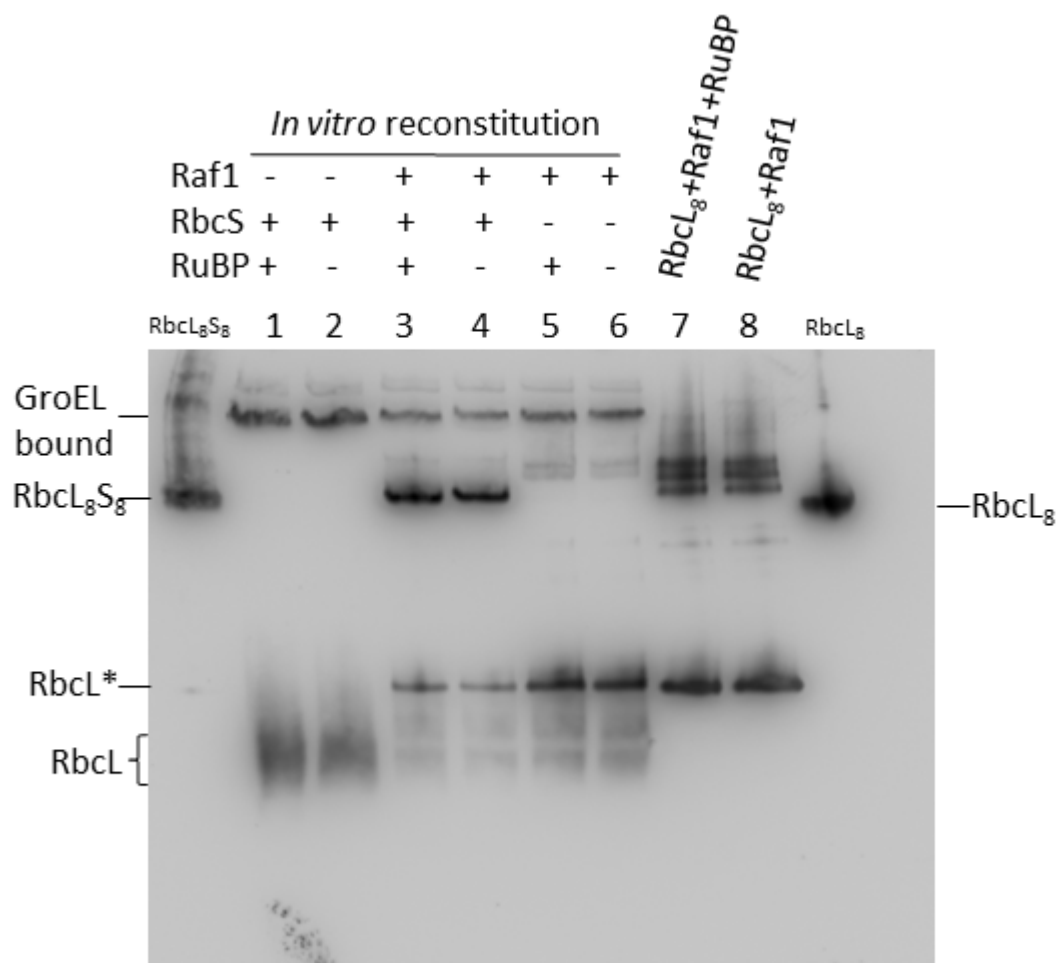


Fig.27: Effect of RuBP on Rubisco assembly and RbcL₈ stability. Analysis of Rubisco reconstitution reactions with homologous components as indicated by native PAGE and immunoblotting with anti-RbcL (lanes 1-6). Complex formation of RbcL₈ with Syn7942-Raf1 and RuBP, analyzed by native PAGE and immunoblotting with anti-RbcL (lanes 7, 8).

3.3 RbcS displaces Raf1 from RbcL₈ at equimolar concentrations

RbcS could efficiently displace Raf1 in Rubisco *in vitro* reconstitution reactions, resulting in RbcL₈RbcS₈ formation and active Rubisco²⁰⁵. However, RbcS was present at tenfold molar excess. To find out at which molar ratio RbcS is able to efficiently displace Raf1 from RbcL₈, pre-formed RbcL₈-Raf1 complexes were incubated with increasing concentrations of RbcS. When Raf1 was added at equimolar concentrations to RbcL₈, RbcL* formation and a HMW species were observed as before (Fig.28, lane 2)²⁰⁵. At equimolar RbcS to RbcL₈-Raf1 complex ratio (Fig.28, left, lane 4,), RbcS already displaced Raf1 quantitatively, indicated by the disappearance of both high HMW RbcL₈-Raf1 and RbcL* complexes. At this ratio, Raf1 was fully released of RbcL₈ and migrated as free Raf1 (Fig.28, right, lane 4). Therefore, Raf1 displacement from RbcL₈ by RbcS binding is highly efficient and essentially irreversible. This might be a result of overlapping binding sites of Raf1 and RbcS on the RbcL₈ core. Higher RbcS binding affinities combined with slower on and off rates would then result in effective

displacement of Raf1 which is a prerequisite for $\text{RbcL}_8\text{RbcS}_8$ formation and enzymatic functionality.

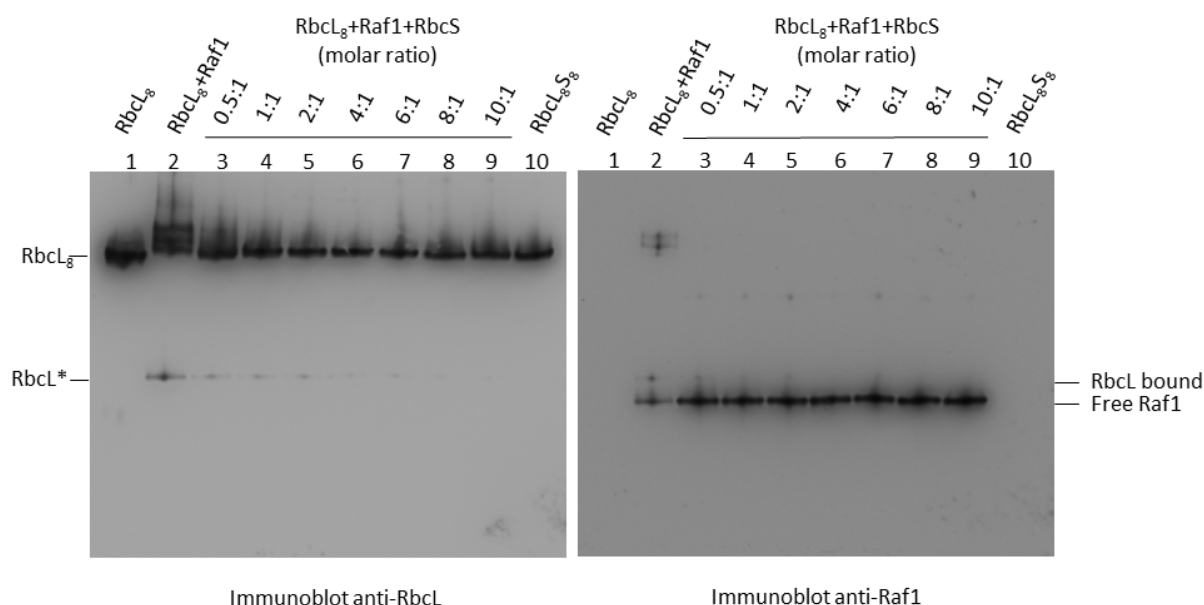


Fig.28: Interaction of RbcS with pre-formed RbcL_8 -Syn7942-Raf1 complexes. $\text{RbcL}_8\text{RbcS}_8$ formation followed by incubation of RbcL_8 -Syn7942-Raf1 (at equimolar concentrations of RbcL to Raf1) with increasing RbcS concentrations as indicated by molar ratios of RbcS to the RbcL_8 -Syn7942-Raf1 complex, analyzed by native PAGE and immunoblotting with anti-Rbcl (left) and anti-Raf1 (right) antibodies.

3.4 Interaction of Raf1 with RbcS

RbcS displaced Raf1 from RbcL_8 very efficiently in the *in vitro* Rubisco reconstitution as well as when Raf1 was bound to purified RbcL_8 . To test whether RbcS binds to Raf1 directly and if Raf1 also plays a direct role in the coordination of RbcS to RbcL_8 , N-terminally FLAG tagged cyanobacterial and plant Raf1 were incubated with their cognate RbcS species. In plants, RbcS is imported in an unfolded state into the chloroplast and presumably interacts with chaperones in the stroma²¹². Therefore, RbcS was added to FLAG-Raf1 in folded and Gdn-HCl denatured unfolded condition. Subsequently an anti-FLAG pull-down was performed and the bound material was analyzed on SDS-PAGE (Fig.29, Fig.30). To rule out background binding of RbcS to FLAG beads, RbcS species were incubated in the absence of FLAG-Raf1 in a separate reaction (Fig.29, Fig.30, lanes 7-12, respectively). BSA was present in the AtRaf1.2 pull-down to minimize potential aggregation of the plant small subunit. BSA is inert and does not bind to the beads. Both FLAG-tagged Syn7942-Raf1 and AtRaf1.2 were effectively pulled down by anti-FLAG beads (Fig.29, Fig.30, lanes 3+6, respectively). RbcS did not bind to either cyanobacterial or plant Raf1 irrespective whether it was denatured or not and eluted in the unbound fractions (Fig.29, Fig.30, lanes 2+5, respectively). AtRbcS was found not to bind to the beads in absence of FLAG-Raf1 regardless if folded or denatured (Fig.29, lanes 9+12). Syn6301-RbcS was found to a small extent in the Syn7942-Raf1 IP fraction when incubated

in denatured state (Fig.29, lane 6), however this was also observed in the absence of FLAG-Raf1, suggesting direct binding to the anti-FLAG beads as observed in the control reaction (Fig.29, lane 12). Therefore, a direct interaction of Raf1 with either native or unfolded RbcS can be ruled out. Importantly, Raf1 was shown to pull-down Rubisco large subunit in a similar co-immunoprecipitation experiment^{205,209}.

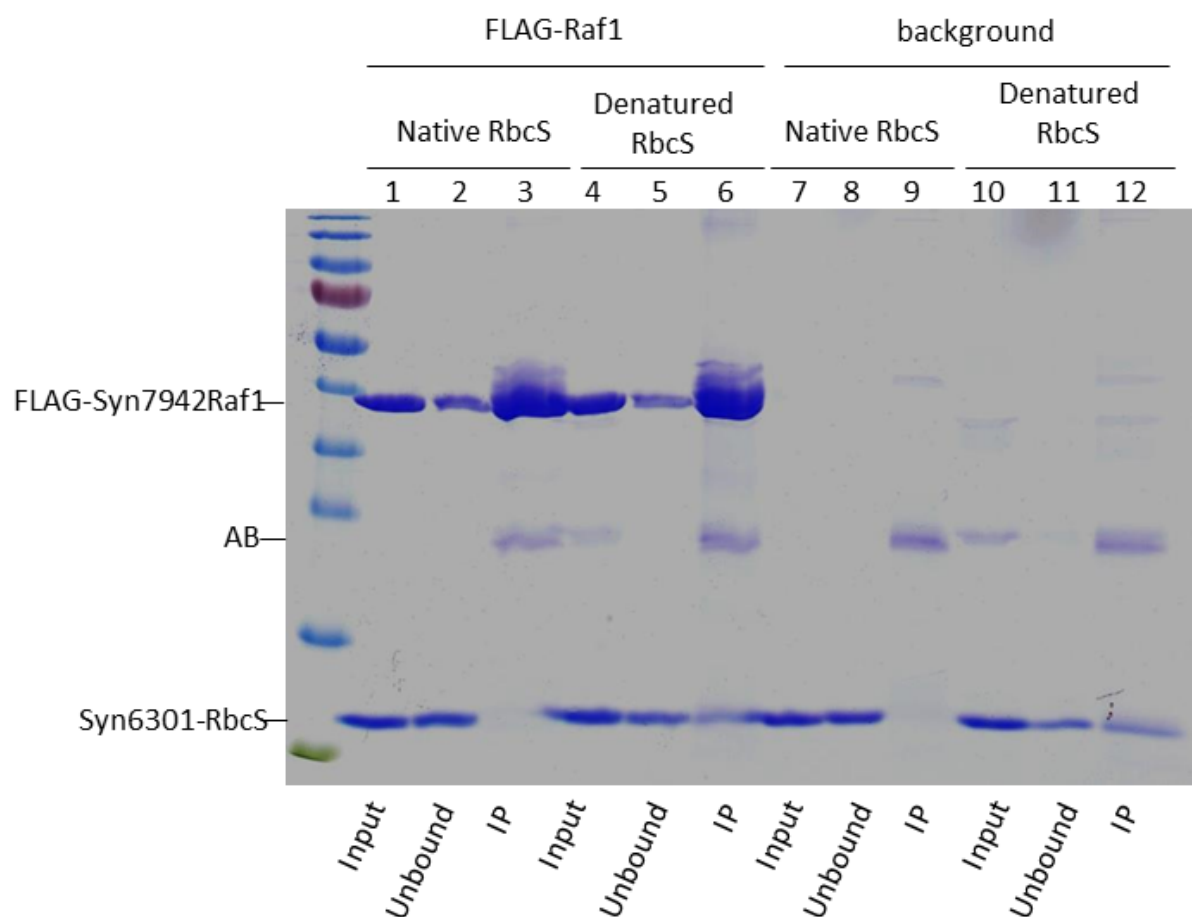


Fig.29: Interaction of Syn6301-RbcS with Syn7942-Raf1. FLAG tagged Syn7942-Raf1 was incubated with either native or denatured Syn6301-RbcS and subsequently an anti-FLAG pull down was performed with anti-FLAG beads and input, unbound (flow-through) and immunoprecipitated (IP) fractions were analyzed on SDS-PAGE followed by Coomassie staining. AB = antibody light chain.

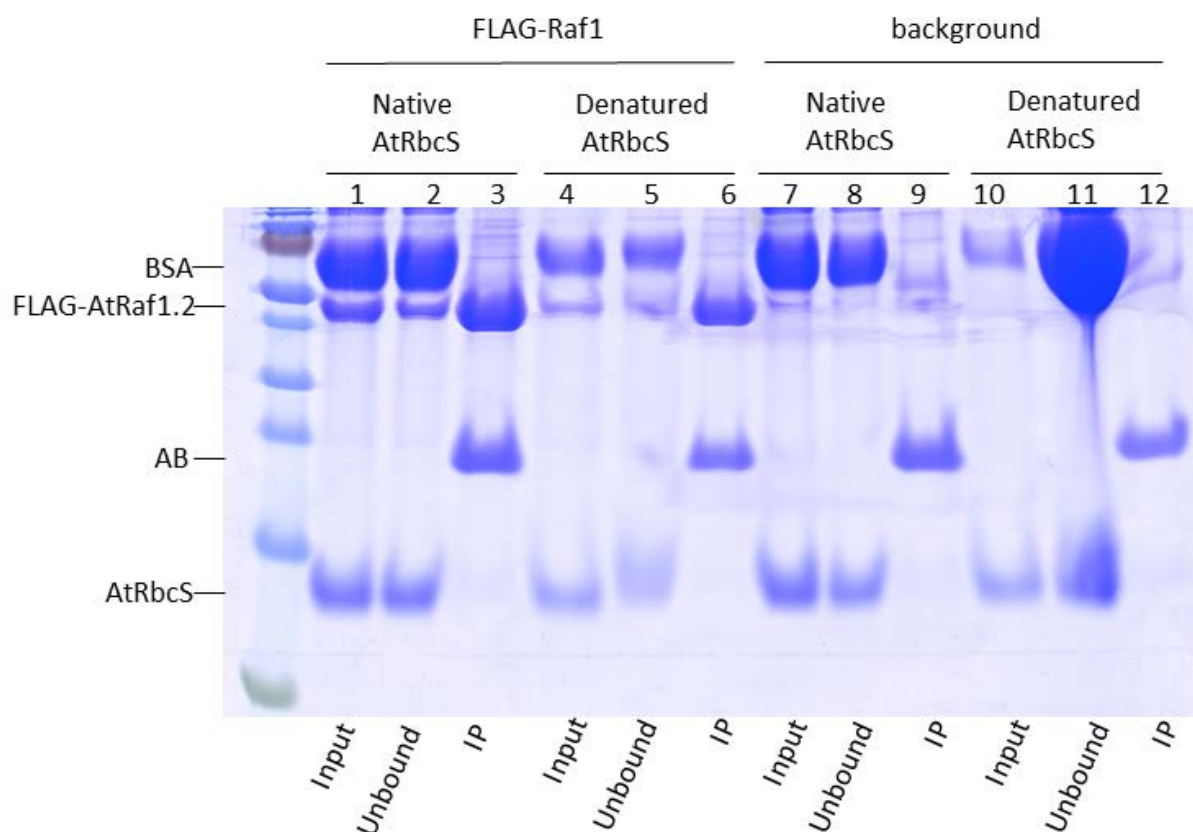


Fig.30: Interaction of At-RbcS with AtRaf1.2. FLAG tagged AtRaf1.2 was incubated with either native or denatured At-RbcS and subsequently an anti-FLAG pull down was performed with anti-FLAG beads and input, unbound (flow-through) and immunoprecipitated (IP) fractions were analyzed on SDS-PAGE followed by Coomassie staining. AB = antibody light chain.

3.5 Interaction of Raf1, Raf2 and RbcS

Raf2 has been identified in some organisms containing form IB Rubisco²¹⁰ and also in organisms containing form IA Rubisco in α -carboxysomes²²⁹. The crystal structure of Raf2 from the chemoautotrophic bacterium *Thiomonas intermedia* K12 showed that the protein forms a homodimer of ~10-kDa subunits with homology to pterin-4 α -carbinolamine dehydratase (PCD) enzymes²²⁹. The overall PCD fold is preserved in Raf2 but the active site cleft present in PCD enzymes is disrupted. Loss of Raf2 function in maize resulted in a phenotype similar to disruption of the *raf1* gene, however the phenotype was less pronounced than in the *raf1* knock-out. Chemical crosslinking followed by immunoprecipitation suggested that Raf2 interacts with imported RbcS subunits and to a lesser extent with RbcL in the chloroplast stroma²¹⁰. It has been proposed that Raf1, Raf2 and Bsd2 form transient complexes with RbcS that maintain RbcS competence for assembly with RbcL²¹⁰. Plant Raf2 contains an additional domain, not present in cyanobacterial Raf2, which may be involved in mediating these interactions. To investigate the potential mechanism of Raf2 action and its putative role in RbcS and/or RbcL binding, pull-down experiments with FLAG-tagged Syn7002-Raf1, Syn7002-Raf2 and Syn7002-RbcS have been performed (Fig.31). FLAG tagged Syn7002-Raf1 was immunoprecipitated effectively by anti FLAG beads (Fig.31, lanes 3+6). Syn7002-Raf2 did not

interact with Syn7002-Raf1 and eluted exclusively in the unbound fraction (Fig.31, lane 2). Furthermore, when Syn7002-Raf1 was incubated with Syn7002-Raf2 and Syn7002-RbcS, neither Syn7002-Raf2 nor Syn7002-RbcS co-immunoprecipitated with Syn7002-Raf1 and eluted in the unbound fraction (Fig.31, lane 5). Therefore, Syn7002-Raf2 does not directly interact with Syn7002-Raf1 and does not mediate RbcS binding to Raf1 to form a ternary complex. Both Syn7002-Raf2 and Syn7002-RbcS did not bind to the anti FLAG beads in the control reactions (Fig.31, lanes 9+12).

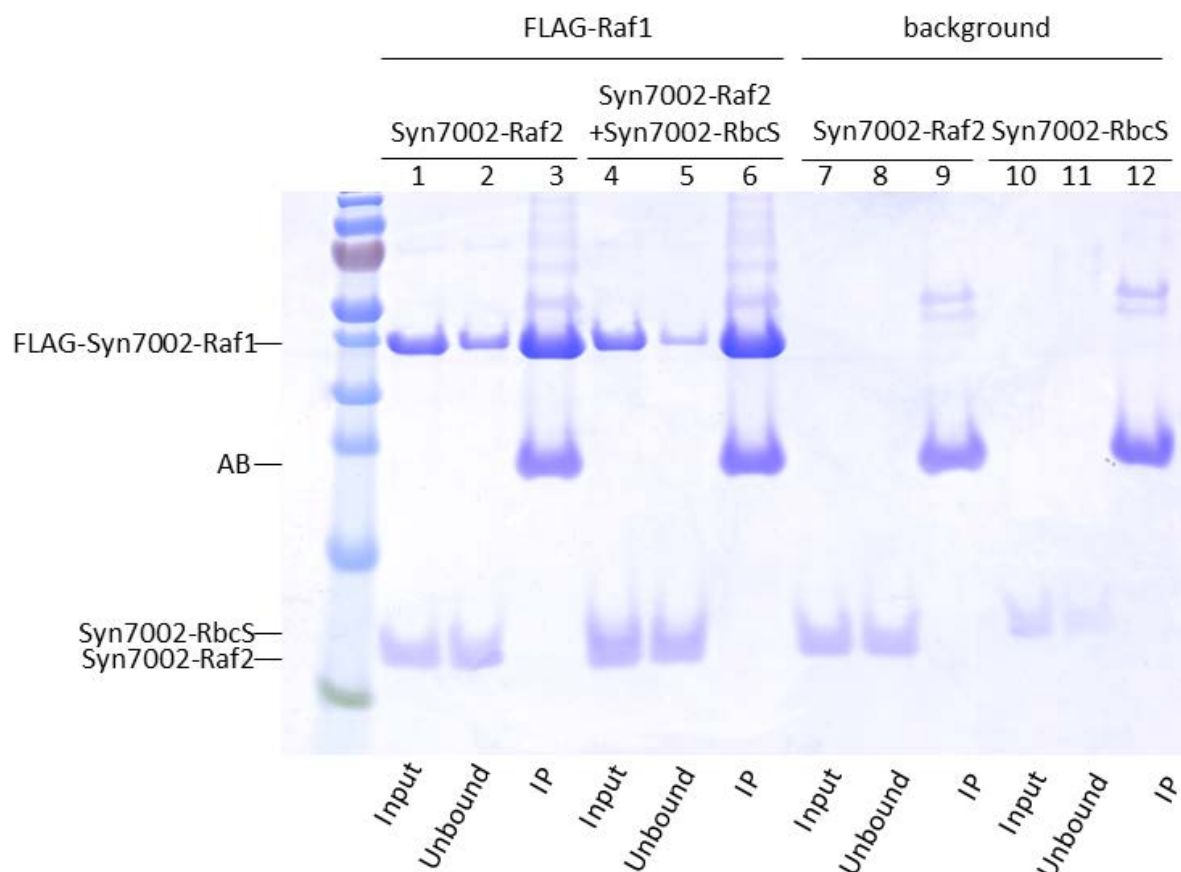
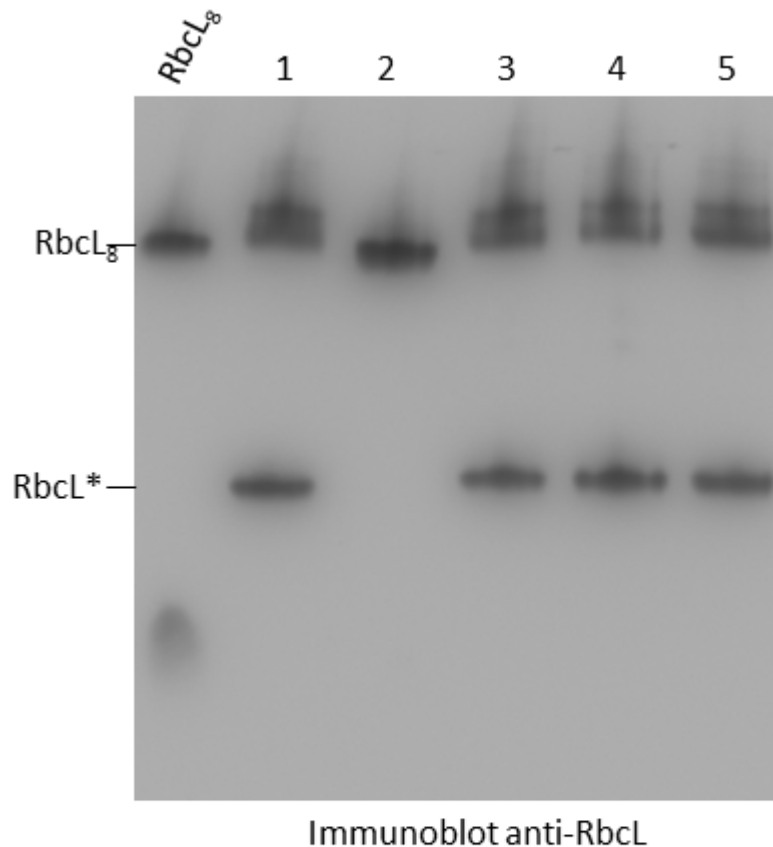


Fig.31: Interaction of Raf1 with Raf2 and RbcS. FLAG tagged Syn7002-Raf1 (top) was incubated with either Syn7002-Raf2 or with Syn7002-Raf2 together with Syn7002-RbcS and subsequently an anti-FLAG pull down was performed with anti-FLAG beads and input, unbound (flow-through) and immunoprecipitated (IP) fractions were analyzed on SDS-PAGE followed by coomassie staining. AB = antibody light chain.

It is important to note that Raf2 homologs from cyanobacterial species containing form IB Rubisco share very low sequence identity with plant Raf2 (below 20% sequence identity) and cyanobacterial form IA Raf2 (20% sequence identity). Furthermore, both Syn7002-Raf2 and Syn6301-Raf2 share the canonical pterin-4a-carbinolamine dehydratase (PCD) catalytic motif [DE]-x(3)-H-H-P-x(5)-[YW]-x(9)-H-x(8)-D²¹¹. These Raf2 verions are the only homologs found in form IB cyanobacteria, suggesting that these proteins rather function as real dehydratases than having a Rubisco assembly associated function in these organisms. In contrast, the acRaf *Thiomonas* homolog of plant Raf2 as well as plant Raf2 show alterations to the canonical

enzyme motif and plants have additional genes of real PCD function²¹¹. Syn7002-Raf2 was chosen over Syn6301-Raf2 since Syn6301-Raf2 is annotated having an additional N-terminal part which renders the enzyme insoluble in *E.coli* (data not shown). To test whether Syn7002-Raf2 could bind to RbcL₈ cores or is able to prevent Raf1 binding, Syn7002-Raf2 was incubated with RbcL₈ in the presence or absence of Syn7002-Raf1 (Fig.32).



1. RbcL₈ + Syn7002-Raf1 (15 min, 25°C)
2. RbcL₈ + Syn7002-Raf2 (15 min, 25°C)
3. RbcL₈ + Syn7002-Raf1 + Syn7002-Raf2 (15 min, 25°C)
4. RbcL₈ + Syn7002-Raf1 (15 min, 25°C) + Syn7002-Raf2
5. RbcL₈ + Syn7002-Raf2 (15 min, 25°C) + Syn7002-Raf1

Fig.32: Interaction of Syn7002-Raf2 with pre-formed RbcL₈ complexes. Complex formation of RbcL₈ with Syn7002-Raf1-and Raf2, analyzed by native PAGE and immunoblotting with anti-RbcL. Samples are purified Syn6301-RbcL₈, incubated with Syn7002-Raf1 and Syn7002-Raf2 as indicated.

Syn7002-Raf2 does neither seem to bind to RbcL₈ cores (Fig.32, lane 2) nor inhibit Syn7002-Raf1 binding to RbcL₈ (Fig.32, lanes 3-5) underlining that form IB cyanobacteria most probably do not possess Raf2 homologs that function in Rubisco assembly.

3.6 *In vitro* reconstitution of plant Rubisco

The first indication that Raf1 is involved in Rubisco assembly was found in *Zea mays*²⁰⁹. It is now clear that Raf1 mediates cyanobacterial form IB Rubisco assembly and its mode of action was described in detail in this study²⁰⁵. *In vitro* reconstitution of plant Rubisco has turned out to be an exceedingly challenging task and any attempts have failed so far. In light of the recent identification of Raf1 as a new factor in plant Rubisco assembly, reconstitution of *Arabidopsis thaliana* Rubisco was attempted. Therefore, the cyanobacterial Rubisco *in vitro* reconstitution system was extended to mimic conditions in plant chloroplasts. Plant factors described to be involved in Rubisco folding and assembly²¹² have been incorporated into the system. Notably, plant chaperonin and co-chaperonin were used for Rubisco folding additionally to the established GroEL/ES system. Furthermore, all factors involved in Rubisco assembly known today were tested in a plant homologous system, specifically Raf1, Raf2, RbcX and RbcS. As described for cyanobacterial Rubisco reconstitution, Gdn-HCl denatured *Arabidopsis thaliana* Rubisco large subunit was diluted in a buffer resembling chloroplast folding conditions (330mM Sorbitol, 50mM Tricine pH 8.3, 100mM KCl, 5mM MgAcetate)¹⁸¹ containing either chloroplast chaperonin 60 with mixed subunits (Cpn60 $\alpha\beta$) or Cpn60 with beta subunits only (Cpn60 β). A mixture of chaperonin 20 (Cpn20) and chaperonin 10 (Cpn10) as co-chaperonin was used. Unfolded proteins like MDH and *Rhodospirillum rubrum* Rubisco have been shown to fold *in vitro* with these components¹⁹³ (Saschenbrecker, unpublished). First, the function of AtRaf1.2 and the heterodimeric AtRaf1.1/2 were tested for their ability to assemble *Arabidopsis* Rubisco (Fig.33). *Arabidopsis* Rubisco large subunit was bound to both AtCpn60 $\alpha\beta$ and AtCpn60 β in the absence of any other factors (Fig.33, left and right, lane 1, respectively). Rubisco reconstitution was unsuccessful under all conditions tested, and no assembly intermediates were observed in any reaction. Even in the presence of Raf1 and RbcS, the large subunit was exclusively detected bound to chaperonin (Fig.33, left and right, lanes 4+8, respectively). Importantly, cyanobacterial Syn6301 Rubisco could be reconstituted with both AtCpn60 $\alpha\beta$ and AtCpn60 β in presence of AtCpn20/10 and Syn7942-Raf1 under these refolding conditions as well as under the standard refolding buffer condition used for GroEL/ES (data not shown). Thus, plant chaperonins are generally able to fold form IB Rubisco as shown before (Windhof, unpublished data).

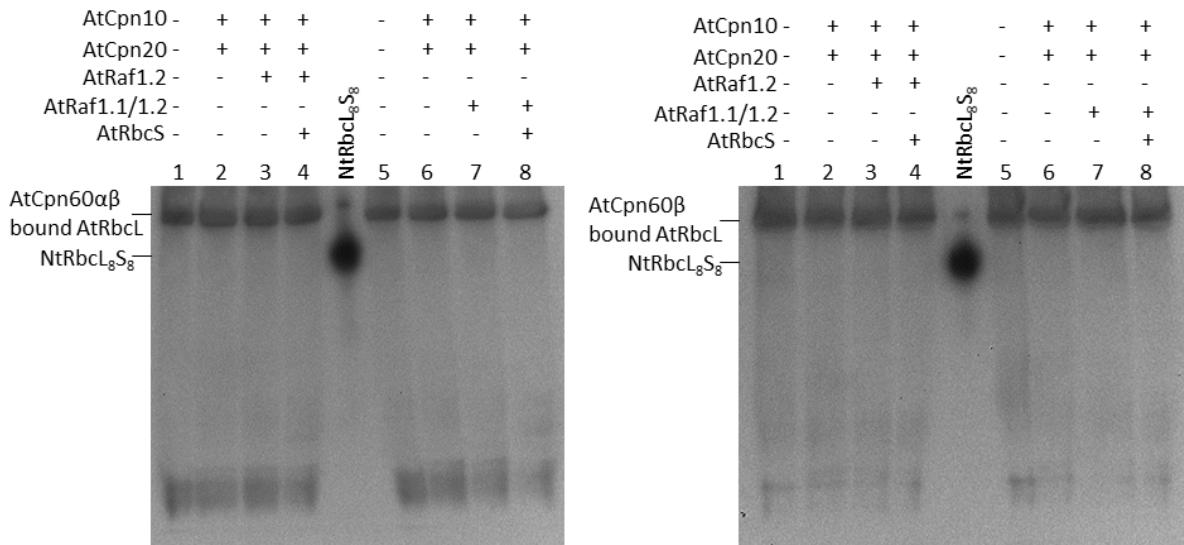


Fig.33: *In vitro* reconstitution attempts of *Arabidopsis thaliana* Rubisco. Analysis of Rubisco reconstitution reactions with homologous components for AtCpn60αβ (left) and AtCpn60β (right), by native PAGE and immunoblotting with anti plant-RbcL. NtRbcL₈S₈ was purified by L. Popilka.

As discussed above, other factors have been described to play a role in Rubisco assembly, e.g. RbcX has been shown to play an essential role in cyanobacterial form IB Rubisco assembly^{201,204} and *Arabidopsis thaliana* RbcX has been implicated to interact with the Rubisco large subunit in co-immunoprecipitation experiments²¹⁴. More recently, Raf2 was identified to interact with the Rubisco small subunit in maize²¹⁰. Therefore, the *in vitro* reconstitution system was extended with *Arabidopsis* RbcXI and RbcXII as well as Raf2 (Fig.33). Standard GroEL/ES buffer conditions were used²⁰⁵ and different combinations of assembly factors were tested. Again, *Arabidopsis* Rubisco large subunit was bound to both GroEL and AtCpn60αβ when only Raf2 was present (Fig.34, left and right, lane 1, respectively). Rubisco reconstitution was unsuccessful in all conditions, and no assembly intermediates were observed. Even in the presence of Raf1/Raf2/RbcXI/RcXII and RbcS, the large subunit was exclusively detected bound to chaperonin (Fig.34, left and right, lanes 5+9, respectively). RuBP was present in the refolding reactions but had no effect on the stabilization on any Rubisco assembly intermediates. Detected bands in the AtCpn60αβ reactions (Fig.34, right, lanes 6-10) are due to cross-reactivity of the anti Syn6301-RbcL antibody which also detects AtCpn20/10 (data not shown).

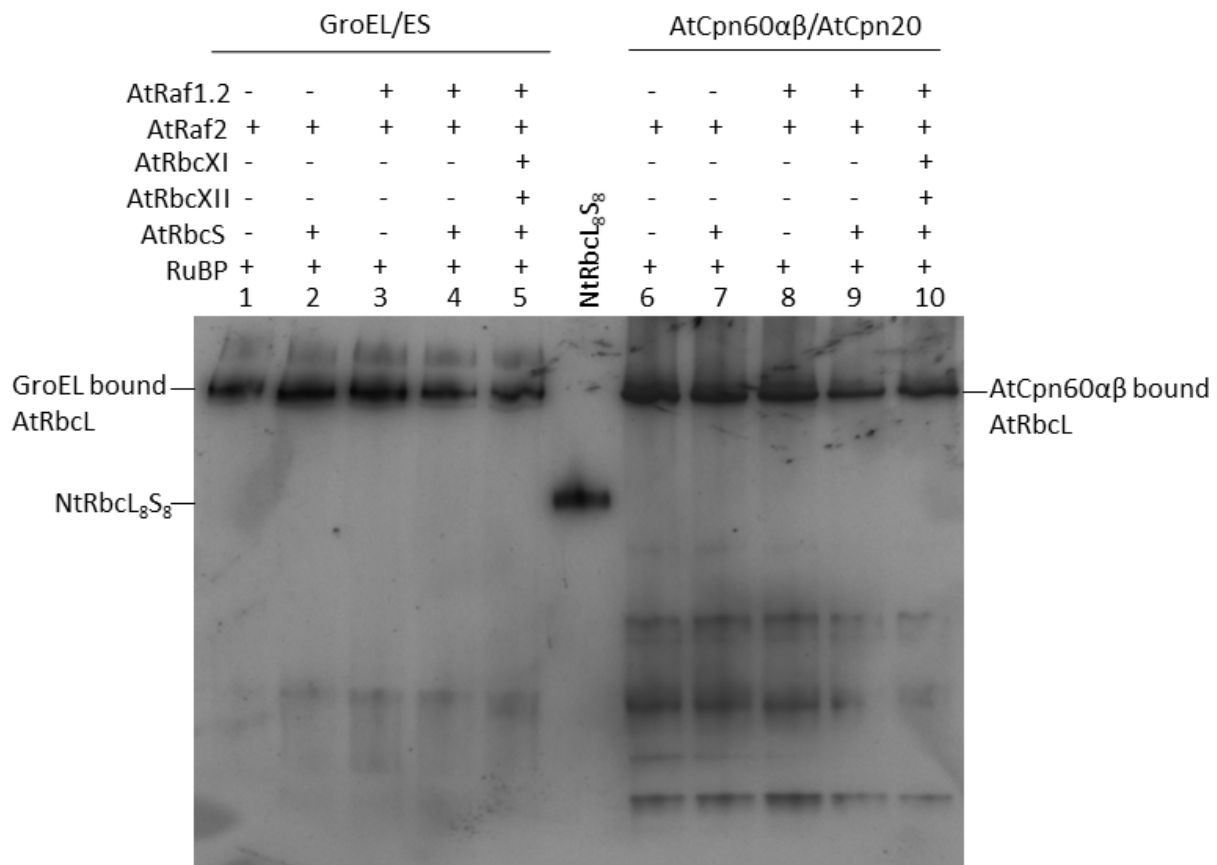


Fig.34: *In vitro* reconstitution of *Arabidopsis thaliana* Rubisco. Analysis of Rubisco reconstitution reactions with homologous components for GroEL/ES (left) and AtCpn60αβ/AtCpn20 (right), by native PAGE and immunoblotting with anti Syn6301-RbcL. NtRbcL₈S₈ was purified by L. Popilka.

3.7 Analysis of eukaryotic RbcX-RbcL interaction

As discussed in Hauser *et al.*²¹², RbcX has been shown to play a role in Rubisco assembly in *E.coli* and is functional in the cyanobacterial Rubisco *in vitro* reconstitution. The potential role of eukaryotic RbcX in the assembly of plant Rubisco, however, is not yet understood. *Arabidopsis thaliana* RbcX was shown to interact with the Rubisco large subunit to some extent in the chloroplast²¹⁴, however a direct role of plant RbcX in Rubisco assembly has not been shown yet. Many eukaryotic photoautotrophs have two RbcX homologs, one that closely resembles the cyanobacterial ortholog, RbcXI, and a more distantly related homolog, RbcXII²⁰¹. The RbcXI and RbcXII from *Arabidopsis thaliana* have been characterized and crystallized, named AtRbcX2 and AtRbcX1, respectively, in these studies^{214,215}. The green algae *Chlamydomonas reinhardtii* contains two RbcXII sequences (CrRbcX-Ila and CrRbcX-Ilb, orthologs of AtRbcXII) and no RbcXI ortholog. Although the RbcX from *Arabidopsis thaliana* is functional when co-expressed with cyanobacterial Rubisco in *E. coli*²¹⁴, a requirement for RbcX in plant Rubisco biogenesis remains to be established. The crystal structures of eukaryotic *Arabidopsis*²¹⁵ and *Chlamydomonas* RbcX²¹⁶ revealed the same structural topology as cyanobacterial RbcX. Each subunit of the dimeric protein consists of four α-helices that form a helical bundle at one end and with a conserved diagonal groove in the center of the dimer.

Structural similarity suggests functional analogy, consequently RbcX-IIa from *Chlamydomonas reinhardtii* and RbcXI and RbcXII from *Arabidopsis thaliana* were tested for their ability to recognize the conserved C-terminal RbcL binding motif of diverse Rubiscos as shown for cyanobacterial RbcX before. RbcX proteins were incubated with a membrane consisting of a peptide array of the last 40 amino acids of RbcL from 20 different species (Fig.35a-c). Corresponding RbcX binding motifs of overlapping binding sites are highlighted in red, as exemplified for CrRbcX-IIa (Fig.35d). Both, *Chlamydomonas* RbcX (Fig.35a) and the two *Arabidopsis* RbcX proteins (Fig.35b, c) bound to the RbcL consensus motif of form IB Rubisco, previously identified for cyanobacterial RbcX, EIKFEF (Fig.35c). The phenylalanine residues in this motif have been shown to play a central role in binding to the RbcX hydrophobic cleft, and large subunit binding was abolished when these residues were mutated²⁰¹. The RbcXII proteins exhibited a very similar binding pattern and were found to bind to more Rubisco species than the RbcXI protein, which might be the result of increased affinity to the RbcL peptides. Intriguingly, while the overall binding pattern between AtRbcXI and AtRbcXII was similar, AtRbcXII displayed seemingly improved recognition of this motif compared to AtRbcXI (Fig.35b, c). Both bound the C-terminal recognition motif of cyanobacterial species, however AtRbcXI did not bind the C-terminal recognition motif of the majority of eukaryotic species, while AtRbcXII did. Specifically, AtRbcXII bound the C-terminal recognition motif of *A. thaliana* while there was no detectable binding to this portion by AtRbcXI (Fig.35b, c, row 13, respectively). Again, this could suggest functional divergence of the two *A. thaliana* RbcX proteins. CrRbcX-IIa binding to eukaryotic form ID peptides might arise from unspecific binding, since these binding patterns include residues FNY which resemble a part of the central form IB binding motif EIKFEF (Fig.35d). Generally, RbcX is not found in species with form ID Rubisco and their respective Rubisco C-termini do not contain the RbcX consensus binding motif²¹². Overall, all three tested eukaryotic RbcX proteins bind to the conserved RbcX binding motif, underlining their functional similarity to cyanobacterial RbcX.

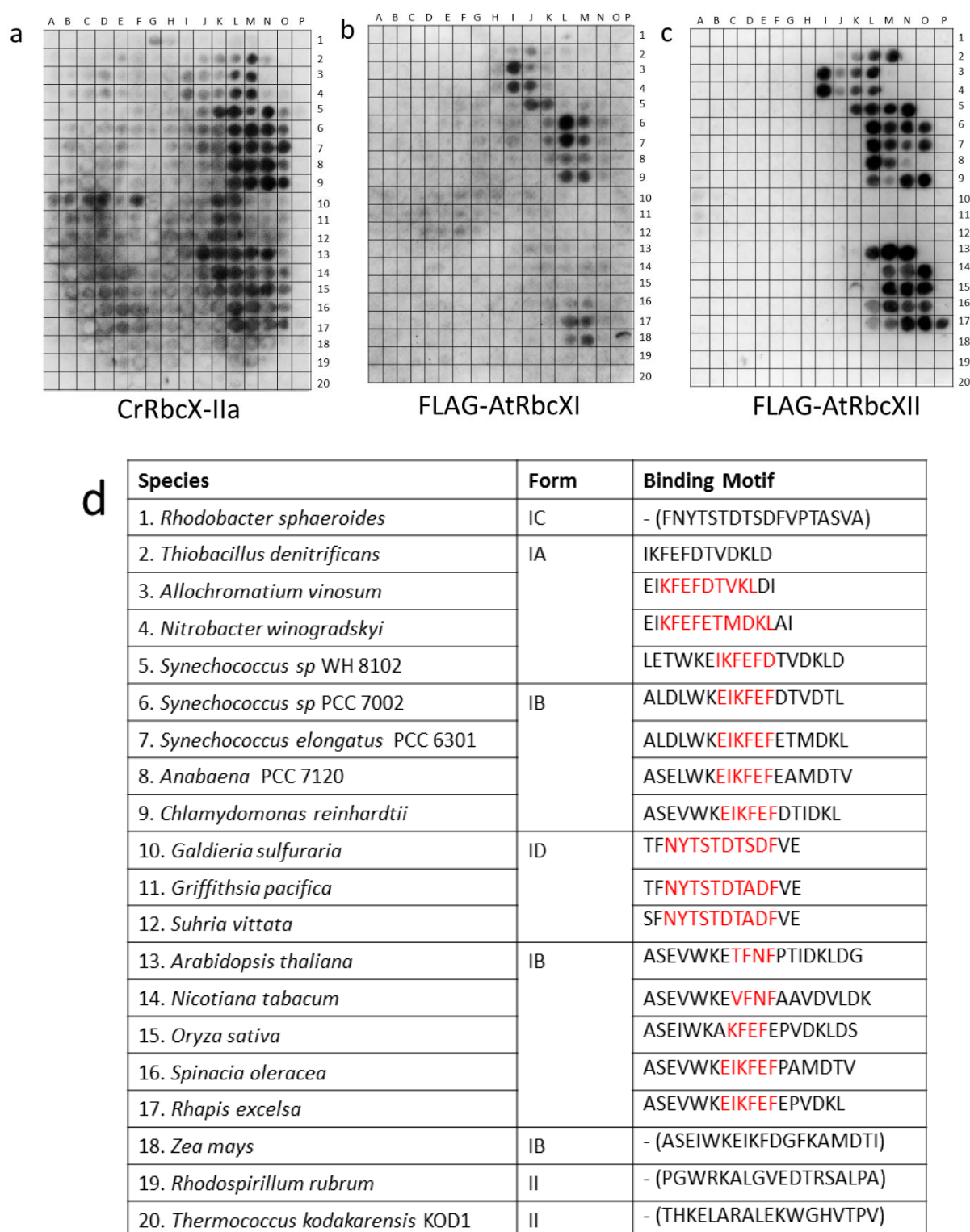


Fig.35: Eukaryotic RbcX binding to RbcL C-terminal peptide membrane. CrRbcX-IIa (a), FLAG-AtRbcXI (b) and FLAG-AtRbcXII (c) were incubated on a membrane consisting of spotted dodecapeptides, with a 10 residue overlap, covering the sequence of the last 40 amino acids of various Rubisco large subunits as listed and analyzed by anti-CrRbcX-IIa and anti-FLAG immunoblot, respectively. Peptides where RbcX was bound are indicated and overlapping binding sites of different peptides for CrRbcX-IIa are highlighted in red (d).

3.8 Functional analysis of eukaryotic RbcX

To investigate the function of eukaryotic RbcX in Rubisco assembly, *in vitro* reconstitution of cyanobacterial Rubisco was performed as described before²⁰⁴. Specifically, CrRbcX-IIa, AtRbcXI, AtRbcXII and functional mutants of these were assessed for their potential to assemble cyanobacterial Syn6301 Rubisco (Fig.36). As observed before, AnaCA-RbcX is able to promote Syn6301 Rubisco assembly²⁰⁴ and due to its high affinity to the Rubisco large subunit stays bound and migrates as an RbcL₈RbcX₈ complex on native PAGE (Fig.36, lane 9)²⁰⁴. Remarkably, all eukaryotic RbcX proteins were able to promote cyanobacterial Rubisco assembly and behaved similarly to AnaCA-RbcX (Fig.36, lanes 3, 5, 7). In all cases, RbcX was bound to the RbcL₈ core and the complexes migrated as HMW RbcL₈RbcX₈ assembly. Mutation of the conserved Arginine residue at position 70 in cyanobacterial RbcX abolished Rubisco assembly and this residue was shown to be important for the formation of a salt bridge with the other Rubisco large subunit in the dimer, promoting antiparallel Rubisco dimer assembly^{201,204}. Interestingly, mutation of the homologous residues in eukaryotic RbcX abolished Rubisco assembly as well (Fig.36, lanes 4, 6, 8).

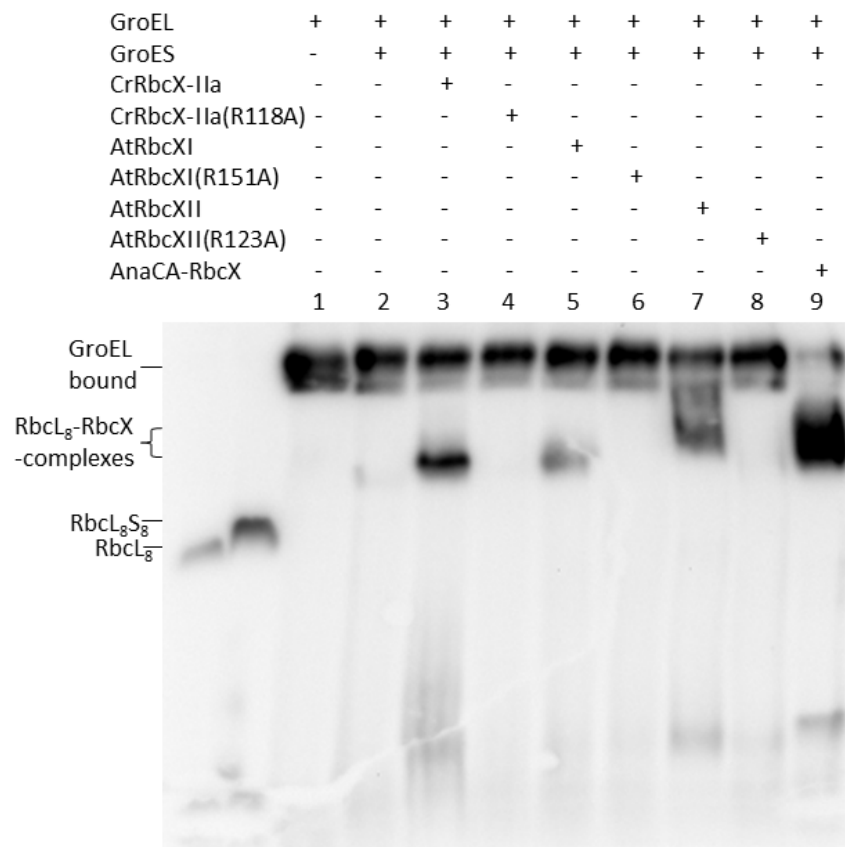


Fig.36: RbcX-mediated assembly of *S. elongatus* Rubisco. Analysis of Rubisco reconstitution reactions in the absence of RbcS as indicated with different RbcX species by native PAGE and immunoblotting with anti-RbcL.

The complexes assembled in the presence of eukaryotic RbcX and were carboxylation active after addition of Syn6301-RbcS to the refolding reactions (Fig.37). In addition CrRbcX-IIb was tested in the refolding reaction, however no Rubisco assembly was observed and no carboxylation activity was detected (Fig.37). Predicted protein sequence shares high homology with CrRbcX-IIa in the core region, however has an elongated C-terminus rich in Alanine and Serine, indicating that it might be wrongly annotated and the real stop codon occurs earlier in the sequence. In the present version, CrRbcX-IIb is unable to promote Rubisco assembly. All eukaryotic RbcX proteins were tested for their ability to promote *Arabidopsis thaliana* Rubisco assembly as tested for Raf1, but none of the RbcX were able to assist in plant Rubisco assembly and neither any assembly products nor carboxylation activity were observed (data not shown).

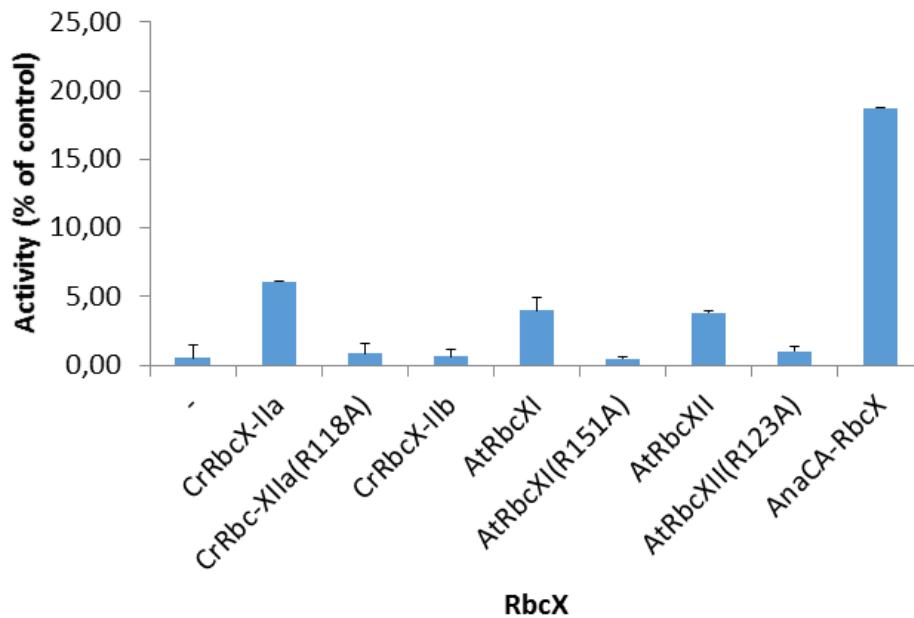


Fig.37: RbcX-mediated assembly of *S. elongatus* Rubisco. Rubisco activity in reactions from Fig.35 after addition of Syn6301-RbcS, which had been absent during reconstitution. Activities are expressed as a percentage of purified RbcL₈ standard supplemented with RbcS (control). Error bars are s.d. ($n = 3$ independent experiments).

3.9 A potential role for RbcX in Rubisco assembly in *Chlamydomonas reinhardtii*

Chlamydomonas reinhardtii is a model organism for studying the effects of genes involved in photosynthesis. It is a single cell organism and compared to plants grows much faster and can be readily genetically modified. It is possible to study the effect of genes involved in photosynthesis even if their function is crucial for cell survival. *Chlamydomonas reinhardtii* can grow on acetate heterotrophically when genetically modified and can be switched to photosynthetic growth under light conditions to assess the functional effect of specific genes on photosynthetic growth. To investigate the effect of RbcX in Rubisco assembly, an RNAi

approach in *Chlamydomonas* was performed. Therefore, a plasmid vector was created with the full sequence of CrRbcX-IIa. The 5'-3' gene was connected via a linker with the 3'-5' reverse sequence of the gene to induce the formation of double stranded hairpin RNA which will be targeted by the cellular RNAi machinery and silence the expression of the respective gene. CrRbcX-IIa was chosen since it was the functional form of the two RbcX copies in *Chlamydomonas* (Fig.36, Fig.37). Nonetheless, CrRbcX-IIa-and IIb are more than 50% identical in the core region of the gene, when the C-terminal extension of CrRbcX-IIb is omitted. Hence, it is very likely that short siRNA fragments created from CrRbcX-IIa will also target CrRbcX-IIb mRNA, since short regions of the gene have even higher genetic identity than the full sequence. *Chlamydomonas* cells were transformed with the silencing vector that also includes a gene for aminoglycoside 3'-phosphotransferase, conferring resistance to paromomycin for antibiotic selection²¹⁷. *Chlamydomonas* transformants were selected on paromomycin containing medium, supplemented with acetate as a carbon source, and then tested for their ability to grow and survive on minimal medium under light (Fig.38a).

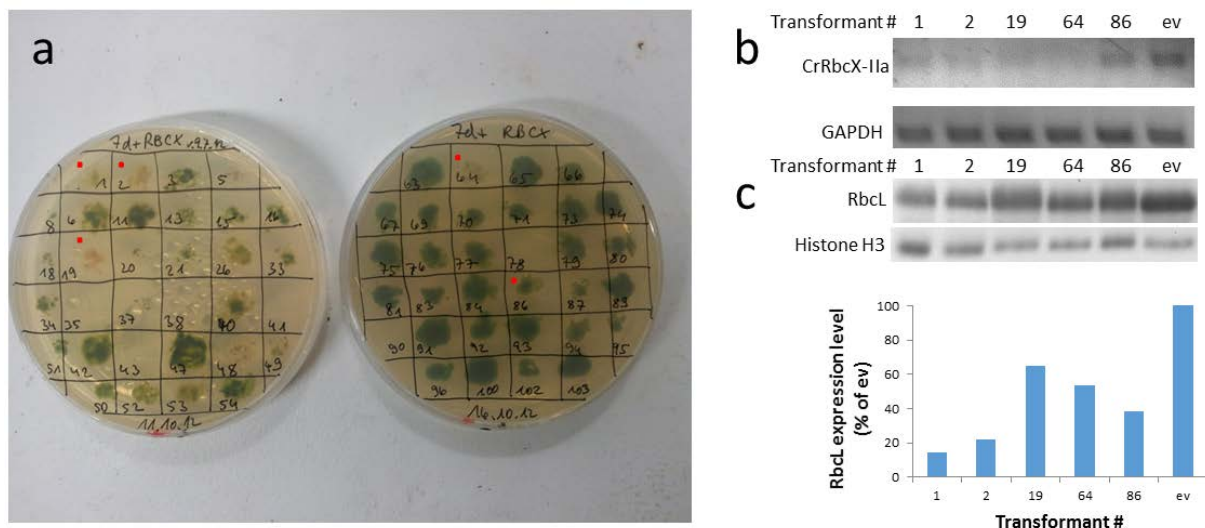


Fig.38: Silencing of RbcX by expression of RbcX inverted repeat transgenes in *Chlamydomonas* transformants. (a) After antibiotic selection on acetate plates, *Chlamydomonas* transformants were grown under light on minimal medium for phenotypic analyses. Transformants that displayed a photosynthetic growth defect, characterized by pale cells, were selected for biochemical analysis (red dots). (b) mRNA levels of selected transformants measured by PCR of CrRbcX-IIa and GAPDH as control, ev = empty vector control. (c) Analysis of RbcL protein level in transformants by SDS-PAGE and immunoblot against RbcL and Histone H3 as control (35 µg loaded each lane). The bar diagram depicts the level of Rubisco large subunit normalized to the internal Histone H3 expression level, ev control was set as 100%.

Most transformants did not show any photosynthetic growth defect, most probably due to inefficient silencing of RbcX. However, several transformants displayed a severe photosynthetic growth defect characterized by a pale appearance of cells (Fig.38a, red dots). These cells were selected for further biochemical analysis (transformants #1, 2, 19, 64, 86) and grown in liquid medium under photosynthetic growth conditions. To assess the efficiency of RbcX silencing, total mRNA was extracted and PCR reactions with primers specific for

CrRbcX-IIa and GAPDH as control were conducted. Intriguingly, all transformants showed reduced CrRbcX-IIa mRNA levels, especially in transformants #1, 2, 19 and 64 CrRbcX-IIa mRNA was barely detectable (Fig.38b). Transformant #86 had a somewhat reduced CrRbcX-IIa mRNA level compared to the empty vector control, where the RbcX inverted repeat was missing. GAPDH mRNA levels were not affected by CrRbcX-IIa silencing. A decrease in mRNA level does not necessarily result in the same reduction of protein expression. Therefore, RbcX protein levels were analyzed by SDS-PAGE and anti-RbcX immunoblot analyses, however no RbcX protein could be detected in either the transformants nor the empty vector control or wild type *Chlamydomonas* cells. It is possible that RbcX is generally only expressed at low levels or was not highly expressed in cells at the point of analysis. A photosynthetic growth defect can result from many different deficiencies of genes involved in photosynthesis. Nonetheless, since silencing of RbcX should have a direct effect on Rubisco assembly, RbcL levels were analyzed in the transformants by SDS-PAGE and anti-RbcL immunoblot analysis (Fig.38c). As normalization control, Histone H3 levels were analyzed which are not expected to change due to CrRbcX-IIa silencing. Remarkably, all transformants showed reduced RbcL levels compared to the empty vector control, with the most drastic reduction in transformants #1 and 2 where RbcL levels were reduced by ~80% (Fig.38c). In other transformants, RbcL levels were reduced between 40% and 60%. Interestingly, transformant #86 displayed lower RbcL levels than #19 and 64, although the CrRbcX-IIa mRNA level was the highest amongst the transformants, highlighting that reduction in mRNA does not completely correspond to reduced protein levels. Overall, there is strong evidence that RbcX plays a role in Rubisco assembly in *Chlamydomonas* and silencing of RbcX resulted in photosynthetic growth defect and directly influenced RbcL levels *in vivo*. Together with the results of the *in vitro* reconstitution and the biochemical analysis of eukaryotic RbcX it is very clear that eukaryotic RbcX also plays an important role in Rubisco assembly as shown before for cyanobacterial RbcX.

3.10 Interplay of Raf1 and RbcX in Rubisco assembly

Both Raf1 and RbcX were structurally and mechanistically studied in this thesis and shown to play an important role in Rubisco assembly. Raf1 promotes Rubisco assembly by bracketing and stabilizing RbcL antiparallel dimers for assembly into RbcL₈ complexes with four Raf1 dimers bound²⁰⁵. RbcX was shown to bind to the conserved Rubisco C-terminus hence also stabilizing the large subunit and promoting dimerization^{201,204}. Considering their different binding mode, it is possible that both Rubisco assembly chaperones act in the same pathway. To test whether Raf1 and RbcX can cooperate in Rubisco assembly, reconstitution experiments with Syn7002-Rubisco, and the *E. coli* GroEL/ES chaperonin were performed testing different combinations of the Syn7002-Rubisco assembly factors (Fig.39).

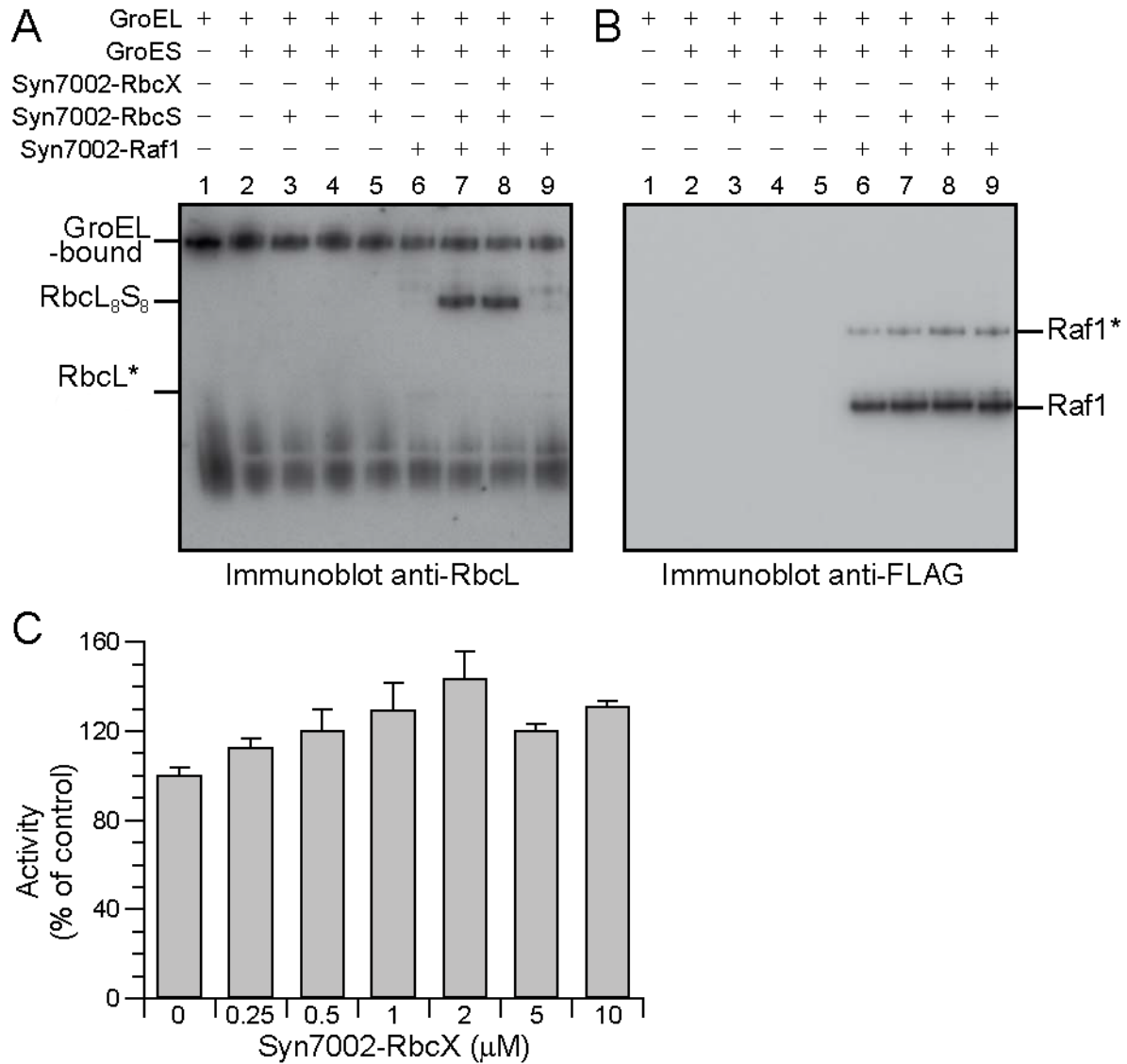


Fig.39: Cooperation of Raf1 and RbcX in Syn7002-Rubisco assembly. A, B: Native-PAGE analysis of Syn7002-Rubisco reconstitution reactions. Denatured Syn7002-RbcL was diluted to a final concentration of 0.5 μ M monomer into buffer containing GroEL (1 μ M), followed by addition of GroES (2 μ M), Syn7002-Raf1-FLAG (0.5 μ M) and Syn7002-RbcX (2 μ M) and Syn7002-RbcS (5 μ M) in the combinations indicated. Refolding was initiated by addition of ATP and reactions were analyzed after incubation for 60 min. at 25 $^{\circ}$ C. Anti-RbcL and Anti-FLAG immunoblots are shown. Raf1* is a higher molecular weight oligomer of Syn7002-Raf1. **C:** Dependence of Rubisco assembly on RbcX concentration in the presence of Raf1. Reconstitution reactions were performed as in lane 8 in (A) at increasing concentrations of Syn7002-RbcX. Rubisco activities obtained after 60 min. are shown. The activity in the absence of Syn7002-RbcX is set to 100%. Error bars represent s.d. values from three independent experiments.

Syn7002 Rubisco fails to reconstitute with chaperonin and RbcX alone, suggesting that an additional factor may be required. Interestingly, when Syn7002-Raf1 was present during refolding of Syn7002-RbcL, very little RbcL* complexes were detectable (Fig.39a, b). This suggested that the interaction of Raf1 with its cognate RbcL is highly dynamic, in contrast to the heterologous interaction between Syn7002-Raf1 and Syn6301-RbcL²⁰⁵. Nevertheless, addition of RbcS gave rise to the formation of Syn7002-RbcL₈S₈ (Fig.39a, b), whereas no

holoenzyme formed with RbcX and RbcS in the absence of Raf1. However, the combination of Raf1 and RbcX resulted in up to ~40% higher enzyme activities, with RbcX showing a dose dependent, saturable effect (Fig.39c). Thus, both assembly factors may cooperate in the same pathway. To further investigate a possible sequential interplay between Raf1 and RbcX, both proteins were added to purified Syn6301-RbcL₈ cores (Fig.40).

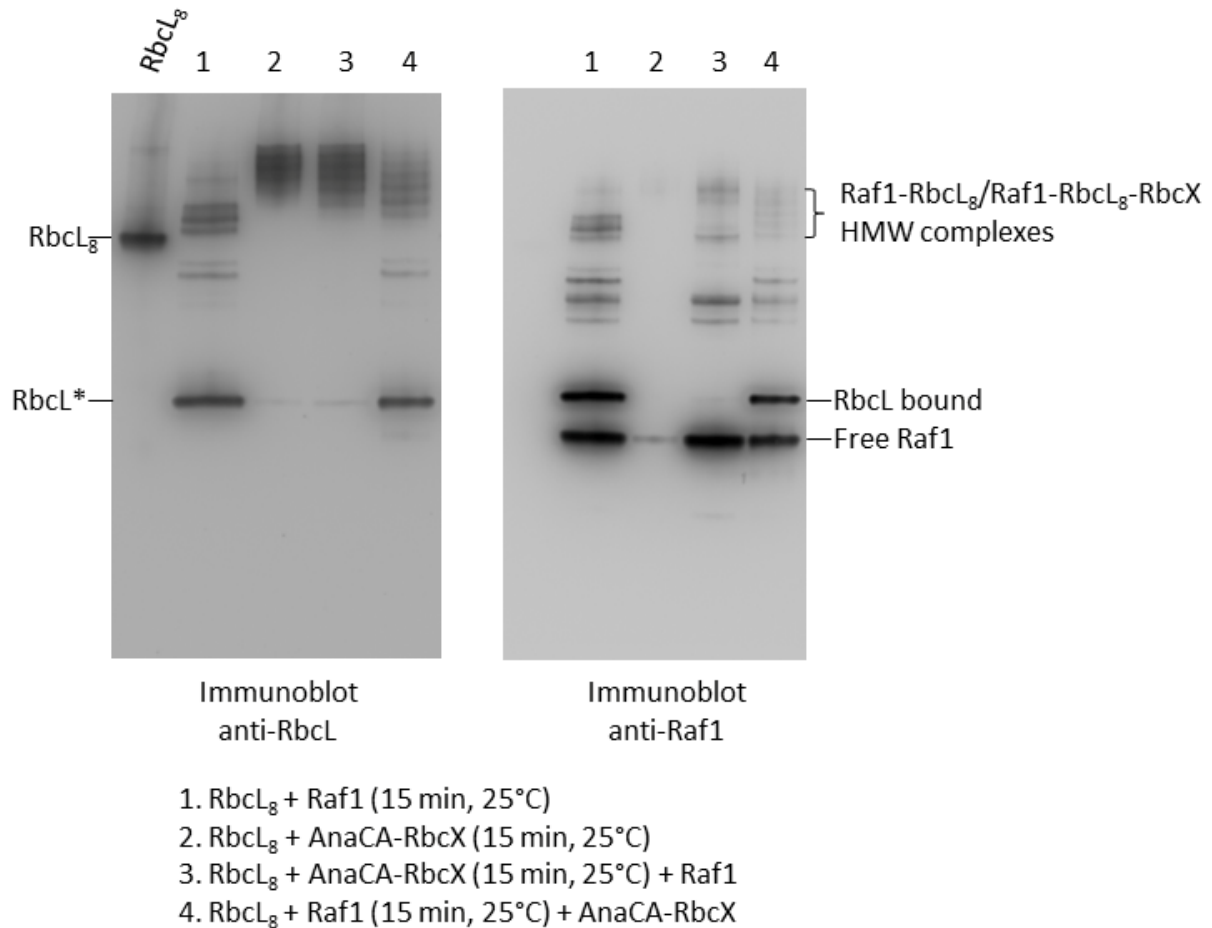


Fig.40: Interaction of Syn7942-Raf1 and AnaCA-RbcX with pre-formed RbcL₈ complexes. Complex formation of RbcL₈ with Syn7942-Raf1 and AnaCA-RbcX, analyzed by native PAGE and immunoblotting with anti-RbcL. Samples are purified RbcL₈, incubated with Syn7942-Raf1 and AnaCA-RbcX as indicated.

Both Syn7942-Raf1 and AnaCA-RbcX were able to bind to RbcL₈ and form HMW complexes, migrating slower than RbcL₈, as observed before (Fig.40, left, lanes 1+2, respectively)^{204,205}. Remarkably, when AnaCA-RbcX was first incubated with RbcL₈ and then Syn7942-Raf1 was added sequentially (Fig.40, left, lane 3), Raf1 was able to shift the HMW complex to a slightly lower HMW form and was detected bound to the RbcL₈ complex migrating slower in native PAGE than when Raf1 was bound alone to RbcL₈ (Fig.40, right, lanes 3+1, respectively). Furthermore, when AnaCA-RbcX was added to pre-formed Syn7942-Raf1-RbcL₈ complexes, again mixed HMW complexes were observed that migrated slower than Raf1 bound to RbcL₈ but faster than RbcX bound to RbcL₈ and Raf1 was detected in these species (Fig.40, left and right, lane 4, respectively). Therefore, it is likely that Raf1 and RbcX can bind simultaneously to RbcL₈ and might act together in Rubisco assembly.

4. Discussion

The mechanism and requirements for the folding and assembly of Rubisco have been studied intensely over the past decades^{190,194,199,201,203,204}. This is a direct consequence of its economic importance. Rubisco has become a paradigm for the study of chaperones and complex assembly in general. Type two Rubisco was used as a GroEL dependent model substrate to study the mechanism of chaperonin action⁴⁹. This simple form of Rubisco found in certain bacteria and dinoflagellates comprises two large subunits, and generally requires only the chaperonin system for folding. However, the evolution of hexadecameric Rubisco, which comprises eight large and eight small subunits, from its dimeric ancestor has rendered Rubisco in most plants, algae, cyanobacteria and proteobacteria dependent on an array of additional factors. These auxiliary factors include several chaperones for assembly as well as ATPases of the AAA+ family for functional maintenance.

Assembly of the RbcL₈RbcS₈ holoenzyme is generally thought to involve the formation of the RbcL₈ core complex followed by the docking of unassembled RbcS subunits, which can fold spontaneously *in vitro*. Various form I cyanobacterial Rubiscos have been expressed in *E. coli*, generally with low yields of 0.1-10%¹⁹⁷, and coexpression of additional GroEL-GroES improves the yield in the case of *Synechococcus elongatus* sp. PCC6301 (Syn6301)^{190,198}. Notably, the Rubiscos from Syn6301 and *Thermosynechococcus elongatus* BP-1 are the only cyanobacterial proteins so far for which a stable RbcL₈ complex could be produced in the absence of RbcS^{200,201,218}. These findings suggested that additional factors, besides the chaperonin, are necessary to ensure the efficient formation of the structurally more complex form I Rubisco.

The cyanobacterial form I Rubisco could be reconstituted *in vitro* using GroEL-GroES and RbcX²⁰⁴. These experiments revealed the requirement of RbcX for stabilization of RbcL in a form that no longer rebinds to GroEL, consistent with findings that RbcX can mediate RbcL₂ and RbcL₈ formation^{201,219}. However, the yield of Rubisco assembly was only 40% owing to the high affinity of the *Anabaena*CA-RbcX to RbcL, which was crucial for the release of large subunits from GroEL, but inhibited RbcS binding to RbcL₈ cores in the heterologous system, suggesting further that an additional factor might be required for Rubisco assembly.

A recent screen of a photosynthetic mutant library in maize identified several mutants with reduced Rubisco content, an important gene found responsible for this phenotype was named Raf1²⁰⁹. In the absence of Raf1, the amount of Rubisco holoenzyme was found to be substantially reduced, although transcription and translation of RbcL and RbcS was normal. Instead, RbcL subunits were associated with the chaperonin, suggesting that Raf1 acts downstream of chaperonin-assisted RbcL folding. Other studies classified Raf1 as a Rubisco

assembly chaperone that may either act in concert with, or possibly in place of, RbcX, however detailed mechanistic insight into Raf1 action in Rubisco assembly was missing^{218,220}.

The present mechanistic and structural study on the recently discovered Rubisco assembly chaperone Raf1, including crystal structures, biochemical analysis, chemical cross-linking and negative-stain EM reconstruction, lead to a new model of chaperone-mediated form I Rubisco assembly that differs significantly from the action of RbcX, the first described Rubisco assembly chaperone. Furthermore, eukaryotic RbcX was investigated to further elucidate its role in Rubisco assembly.

4.1 A new model for Raf1-mediated Rubisco assembly

Assembly of oligomeric protein complexes is widely considered a spontaneous process, and relatively little is known about the machineries that support the formation of specific multiprotein complexes. The biogenesis of hexadecameric Rubisco has emerged as a paradigm of assisted assembly^{221,222}. In this study, plant and cyanobacterial Raf1 proteins were characterized functionally and structurally. The crystal structures of the *A. thaliana* Raf1 domains were solved and the interaction of Raf1 with RbcL was analyzed by multiple biochemical and biophysical approaches. The major intermediate populated during assembly is a complex in which dimeric Raf1 brackets the antiparallel RbcL dimer, thus preventing rebinding of the structurally labile RbcL to GroEL. It is also possible that Raf1 brings RbcL subunits into proximity, thereby promoting dimer formation. The RbcL₂-Raf1 complex is competent for assembly into higher oligomeric states, with the RbcL₈-Raf1₄ complex as the endpoint in absence of RbcS. Finally, binding of RbcS displaces Raf1 and completes assembly of the holoenzyme²⁰⁵. Like the structurally unrelated Rubisco-assembly chaperone RbcX^{217,223}, Raf1 is dimeric and engages in bivalent interactions with RbcL, a principle that probably relates to the antiparallel RbcL dimer being the building block of the RbcL₈ complex. Raf1 consists of an N-terminal α -domain, a flexible linker segment and a C-terminal β -sheet domain that mediates dimerization. Both domains expose conserved interaction surfaces. The α -domains mediate the majority of functionally important contacts with RbcL by bracketing each RbcL dimer at the top and bottom, and the β -domain dimers are positioned coaxially in front of the RbcL₂ units. Because the α -domain alone is essentially inactive, dimerization is critical for Raf1 to achieve the necessary avidity for complex formation with RbcL assembly intermediates. At the same time, the interaction must remain dynamic in order to allow RbcS binding, because the Raf1 α -domain and RbcS have overlapping binding sites on RbcL.

4.2 The role of eukaryotic RbcX in Rubisco assembly

The function of RbcX in cyanobacterial Rubisco assembly has been established and RbcX was the first discovered form I Rubisco assembly chaperone. Coexpression of RbcX with RbcL

and RbcS in *E. coli* enhanced the production of enzymatically active Rubisco²⁰¹. Partial inactivation of the *rbcX* gene in *Synechococcus* sp. PCC7002 resulted in a substantial reduction in Rubisco solubility and activity²⁰⁰, while in *S. elongatus* sp. PCC7942 the *rbcX* gene was shown to be nonessential¹⁹⁸. RbcX is highly conserved in all prokaryotes and eukaryotes containing form IB Rubisco. Structural analysis showed that RbcX is a dimeric, α -helical protein of ~15 kDa subunits^{201,215}. Although the RbcX from *Arabidopsis thaliana* is functional when substituted into *E. coli* expressing cyanobacterial Rubisco, a requirement of RbcX in Rubisco biogenesis in plants remains to be established, although a direct interaction between RbcL and RbcX was detected in chloroplasts²¹⁴. The function of RbcX in cyanobacterial Rubisco assembly was studied in detail and a model has been proposed on the action of RbcX mediated Rubisco assembly²⁰³: after folding and release by GroEL/ES, the flexible C-terminus of folded RbcL binds within the central cleft of RbcX; charge pairs in the RbcL dimer interface and between RbcL and RbcX then mediate the proper antiparallel alignment of the RbcL subunits with RbcX functioning as a molecular clamp; stabilization of the RbcL dimer then shifts the equilibrium towards the formation of the RbcL₈RbcX₈ assembly intermediate; finally, RbcS binding between the RbcL dimers triggers conformational changes in the RbcL subunits, resulting in the displacement of RbcX and formation of the hexadecameric Rubisco²⁰⁴. In this study, *Chlamydomonas reinhardtii* CrRbcX-IIa was characterized biochemically and structurally²¹⁶. The crystal structures of CrRbcX-IIa alone and in complex with the C-terminal peptide of RbcL showed that CrRbcX-IIa shares the structural topology with cyanobacterial and plant RbcX homologs. However, the RbcL peptide bound to CrRbcX-IIa only occupies part of the central hydrophobic cleft of the RbcX dimer, in contrast to the structure of the cyanobacterial RbcX-peptide complex. Nevertheless, CrRbcX-IIa supports the assembly of cyanobacterial Rubisco, although with reduced efficiency compared to cyanobacterial RbcX. Therefore, RbcXII from the green algae *C. reinhardtii* functions as a bona fide Rubisco assembly chaperone, despite its considerable evolutionary distance from cyanobacterial and eukaryotic RbcXI proteins. CrRbcX-IIb was not active in Rubisco assembly, however as discussed in (3.8.) this is presumably owed to incorrect annotation of the coding sequence, since the two CrRbcX homologs display great homology in the core region of the protein. Therefore, it is possible that *Chlamydomonas* has two functional RbcXII homologs that might have arisen from a gene duplication event.

RbcX homologs from *Arabidopsis thaliana* were shown to exhibit similar activity to CrRbcX-IIa in reconstituting cyanobacterial Rubisco *in vitro*. Therefore, all eukaryotic RbcX homologs investigated in this study assist in Rubisco assembly in a similar way to their evolutionary distant cyanobacterial homologs. Interestingly, all eukaryotic RbcX were observed to bind to the conserved RbcL C-terminal motif as shown for cyanobacterial RbcX. Importantly, when the conserved Arginine on the peripheral site was mutated, Rubisco assembly was abolished,

strengthening functional similarity between these RbcX homologs. Furthermore, the effect of RbcX mediated Rubisco assembly *in vivo* was investigated by silencing the RbcX genes in *Chlamydomonas*. Remarkably, in transformants showing a photosynthetic growth defect, RbcX mRNA levels were strongly decreased. In addition, RbcL protein levels were diminished as well in these transformants when compared to control cells. This is the first indication that RbcX plays a functional role in Rubisco assembly in eukaryotes, similarly to the effect of deleting Raf1 in maize. However, the finding that RbcX could not be detected in wild-type cells warrants further analysis of this photosynthetic growth defect. It would now be important to show a direct interaction of RbcX with Rubisco in wild-type cells. Furthermore, microscopic studies on the pyrenoid formation in RbcX-deficient transformants could underline the role of RbcX in Rubisco assembly, since Rubisco defect strains have been shown to lack pyrenoids²²⁴.

4.3 Functional interplay between Raf1 and RbcX in Rubisco assembly

Interestingly, the structurally distinct Raf1 and RbcX assembly proteins perform the same function, but they do so by using different interaction sites on RbcL, that only partially overlap (Fig.41). The boomerang-shaped RbcX dimer binds the C-terminal peptide tail of one RbcL subunit in a central cleft and contacts the N-terminal domain of the adjacent RbcL subunit via a peripheral region^{201,204}. In contrast, the Raf1-RbcL interaction surface appears to be more extensive (Fig.41). An interesting question therefore concerns whether Raf1 and RbcX act in parallel assembly pathways or functionally cooperate *in vivo* to achieve efficient assembly at a biologically relevant timescale. Taking into consideration that the binding sites for RbcX and RbcS have little if any overlap (Fig.41), the previously described RbcL₈-RbcX₈ complex^{203,204} may be an additional assembly intermediate, especially when insufficient RbcS is present. A cooperation of Raf1 and RbcX, rather than mere functional redundancy, would be consistent with the strict co-occurrence of the two factors in photosynthetic organisms.

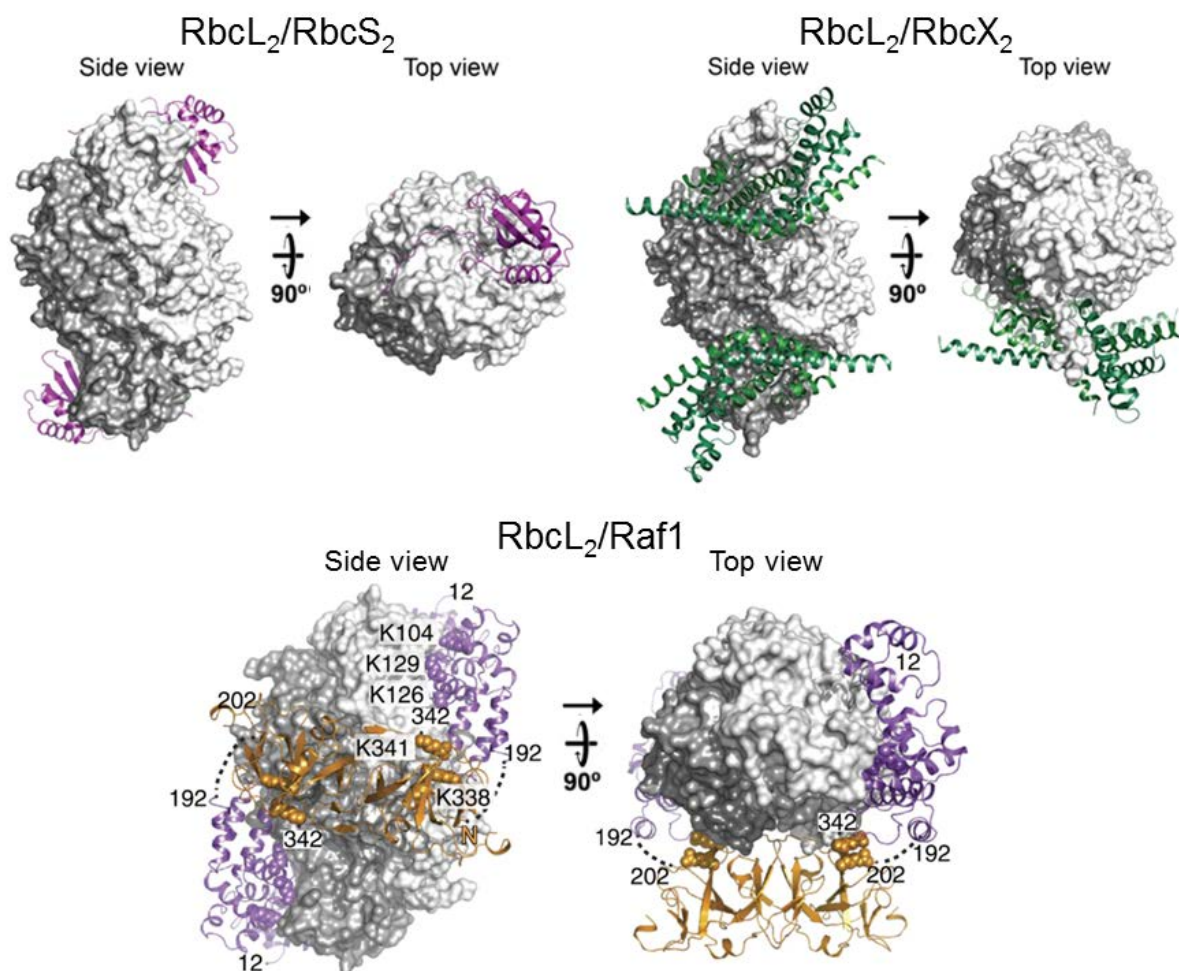


Fig.41: Interaction regions of RbcS, RbcX and Raf1 on RbcL₂. Top, left: RbcL₂RbcS₂ complex extracted from the Rubisco holoenzyme structure (PDB 1RBL) and magnified. Left, side-view; right, top-view. The RbcL subunits are shown in surface representation; RbcS in ribbon representation (purple). Top, right: RbcL₂RbcX₂ complex extracted from the RbcL₈RbcX₈ structure (PDB 3RG6). The RbcX dimers are shown in dark green. The conspicuous protrusion in RbcL is the extended C-terminal peptide which binds into the central cleft of RbcX. Bottom: RbcL₂Raf1 protomer of the RbcL₈Raf1₄ complex²⁰⁵, based on the cross-linking data (cross-linked residues are indicated). Functionally critical Raf1 residues are shown in space-filling representation. Raf1 α -domain is depicted in purple, Raf1 β -domain in orange. The inter-domain linker region is indicated as a dotted line.

The chaperone requirement for the folding and assembly of hexadecameric (form I) Rubisco is remarkably complex, thus raising the question of why such an important and abundant enzyme has failed to evolve a more robust biogenesis pathway. Form I Rubisco evolved from the simpler dimeric form II enzyme about 2.5 billion years ago to adapt to the increasing levels of oxygen²²⁵. Both the form I and form II RbcL subunits require the GroEL-ES chaperonin for folding, like many other TIM-barrel proteins^{48,226}. However, only the RbcL of form I Rubisco remains structurally labile after folding and tends to rebind to chaperonin rather than to spontaneously assemble. Therefore, it is possible that Rubisco assembly chaperones function firstly in stabilizing the large subunit and preventing it from rebinding to GroEL. Interestingly, GroEL has overlapping binding sites with Raf1 on RbcL when tested on the Syn6301-RbcL peptide membrane (see 3.1). Substrate binding to GroEL was shown to involve specific peptide regions with high hydrophobicity as found for the model GroEL binding protein GroES

(GGIVLTG)²²⁷. Intriguingly, the GroES peptide region that binds to GroEL is part of a mobile loop and it was observed that GroES mobile loop-like hydrophobic patches in GroEL substrates are present in the unstructured region of proteins, specifically in loop or β -sheeted regions²²⁷. Remarkably, the Syn6301-RbcL peptide recognized by GroEL is part of a loop region in the TIM barrel domain between a β -sheet and α -helix (Fig.42). The detected binding motif is rather hydrophobic (LGSTIKPKLGL) and shows physico-chemical similarity to a peptide motif in α -lactalbumin which was predicted to be responsible for GroEL binding (KILDIKGI)²²⁷. Interestingly, the GroEL binding region is in close proximity to one of the binding patches of Raf1 which was also found in cross-linking experiments (Fig.42)²⁰⁵. The observed GroEL binding region on the folded Syn6301-RbcL (Fig.42, left) is solvent accessible and might be the reason for facile re-association of the folded large subunit with GroEL as observed in the absence of assembly chaperones^{204,205}. This region becomes protected in the RbcL dimer and Raf1 or RbcX binding most likely further restrict accessibility of this loop region. Therefore, Raf1 and RbcX mediated dimerization potentially prevents this region from binding back to GroEL.

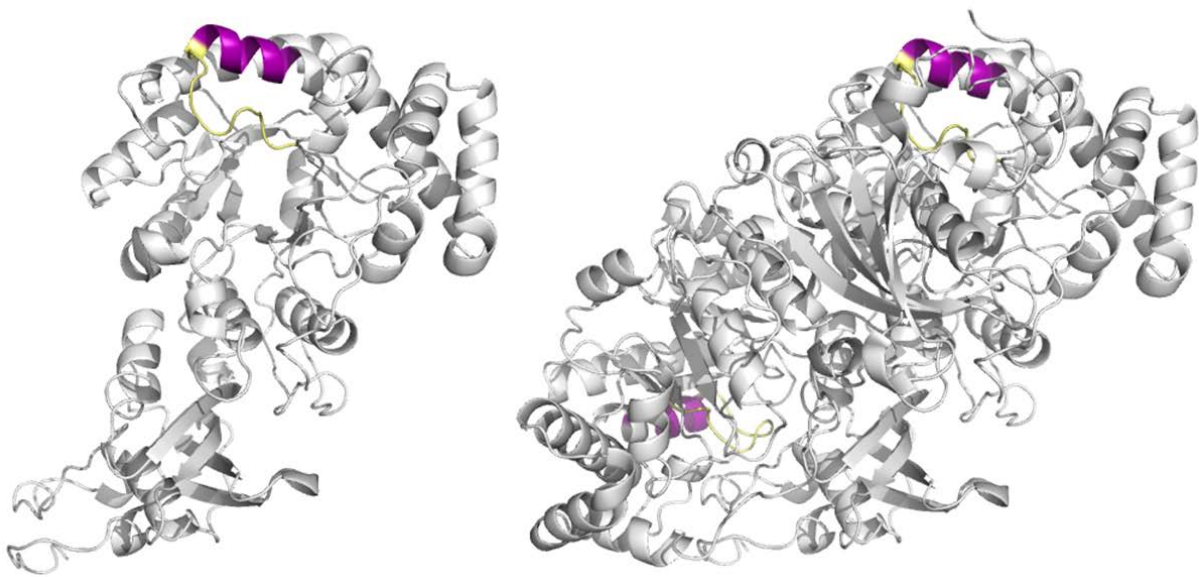


Fig.42: GroEL binding to Syn6301-RbcL. Left: GroEL binding region based on the Syn6301-RbcL monomer (white) highlighted in pale yellow. Adjacent Syn7942-Raf1 binding region is shown in purple (based on binding to peptide membrane (Fig.22, left; Fig.19, left)). Right: Same orientation as in left, depicted for the Syn6301-RbcL dimer, similar color coding.

In the absence of assembly factors in the *in vitro* system the kinetics of dimerization might be too slow and thus RbcL does not assemble spontaneously and rather is detected bound to GroEL. Interestingly, when Syn6301 RbcL is expressed in *E.coli*, limited amounts of RbcL₈ are formed. Other chaperones like Hsp70/40 might bind to the large subunit and high expression of large subunits might allow spontaneous dimerization to a certain extent. However, most of the protein still aggregates, indicating that Raf1 and RbcX play further active and essential roles in the assembly of Rubisco rather than simply preventing binding of folded RbcL to

GroEL. Remarkably, in case of Syn7002 Rubisco, assembly was completely dependent on Raf1. However, when the influence of the RbcX level present was tested, up to 40% higher yields of assembled Rubisco were observed. Syn7002 Rubisco could not be reconstituted with RbcX alone and does not form RbcL₈ cores when expressed in *E.coli*, but assembles when co-expressed with RbcX²⁰¹. This represents a model system to study the potential interplay between Raf1 and RbcX. Coexpression experiments of Syn7002 Rubisco alone and with different combinations of the assembly chaperones in *E.coli* should be undertaken to unravel this fundamental question in Rubisco assembly.

4.4 *In vitro* reconstitution of plant Rubisco

Given that the RbcL subunits of all form I Rubiscos are highly homologous, it is surprising that the Rubisco of higher plants has so far been refractory to reconstitution in *E. coli* or *in vitro*. In the past it was suggested that maybe additional factors are required for the assembly of the plant enzyme²¹². The attempt to *in vitro* reconstitute *Arabidopsis thaliana* Rubisco in this study is perhaps the most comprehensive effort that has been undertaken until today (see section 3.6). In this system, almost all previously and recently described Rubisco assembly chaperones were present. Furthermore, the proteins used in the assembly reaction are all from *Arabidopsis*, thus representing a purely homologous system. Both chaperonins used, GroEL/ES and AtCpn60 $\alpha\beta$ /AtCpn60 β /Cpn20/Cpn10, were shown to be active in folding cyanobacterial form I Rubisco large subunits. However, neither Rubisco assembly nor any assembly intermediates were observed when reactions were analyzed on native PAGE immunoblot experiments. Instead, plant Rubisco large subunit was exclusively detected bound to the chaperonin systems. One possible explanation for this could be insufficient folding of the large subunit. The bacterial chaperonin may not be able to fold the large subunit to its native state rather than forming a compact intermediate that might be not recognized by the assembly chaperones. However, this would not explain why the *Arabidopsis* chaperonin system should be unable to fold its cognate Rubisco large subunit. In this study, both an established GroEL/ES refolding buffer system was used and a buffer system, in which isolated pea chloroplasts were able to translate and assemble Rubisco *de novo*¹⁸¹. It is still conceivable that factors were present in the isolated chloroplast that were missing in the pure *in vitro* system, i.e. sugar substrates or uncharacterized assembly chaperones. Clearly, the large subunit was able to bind to AtCpn60, however it was not determined if it was properly folded. The presence of its substrate RuBP did not show an effect in stabilizing any assembly intermediates (see 3.2 and 3.6). Moreover, recently it was shown that for red-type prokaryotic Rubisco, the small subunit was sufficient for Rubisco assembly²⁰⁵. These subunits have an additional C-terminal loop that was required for Rubisco assembly. In contrast to green-type bacterial form I RbcL however, the large subunits were not bound to GroEL after folding in the absence of RbcS,

Discussion

indicating that these are less prone to rebind to GroEL and probably dimerize more readily. Furthermore, any described green-type Rubisco assembly chaperone is not found by homology search in red-type prokaryotic or eukaryotic photosynthetic organisms, suggesting that Rubisco assembly might have evolved in different ways in these organisms. Nevertheless, eukaryotic red-type Rubisco was resistant to *in vitro* reconstitution in the presence of cognate RbcS, and these subunits again were solely detected bound to GroEL (Jidnyasa Joshi, unpublished). Possibly, similar to plant Rubisco, other eukaryotic Rubiscos might require their specific chaperonin environment. Plant RbcS does not comprise the C-terminal extension and was not able to assemble red-type eukaryotic Rubisco. In plants, RbcS is located in the nucleus and imported into the chloroplast⁸². Therefore, in chloroplasts Rubisco assembly might be timely and spatially regulated in a way that Rubisco assembly is achieved at the RbcS import sites next to the chloroplast membrane. Indeed, recent findings suggest such a coupled mechanism for *Chlamydomonas*²²⁸. In such a situation, RbcX might be crucial to stabilize RbcL₈ cores until RbcS is available for holoenzyme assembly. Raf1 might readily bind to newly synthesized and folded large subunits but might not be able to stabilize RbcL₈ until the small subunits become available. In order to test RbcS import conditions, in some of the plant reconstitution experiments presented in this study, denatured RbcS was added to fold spontaneously as would occur after import into the chloroplast to rule out the possibility that RbcS must fold in the presence of assembled RbcL. However, Rubisco assembly was still unsuccessful even in additional presence of assembly chaperones (as in 3.6, data not shown). The most recently described potential Rubisco assembly factor Raf2 was also tested for its effect on *in vitro* Rubisco assembly. Chemical crosslinking followed by immunoprecipitation suggested that Raf2 interacts with imported RbcS subunits and to a lesser extent with RbcL in the chloroplast stroma²¹⁰, but Raf2 was not able to bind to the large subunit in the *in vitro* reconstitution system and in combination with other chaperones did not promote Rubisco assembly. Furthermore, cyanobacterial Syn7002-Raf2 did neither exhibit stable interactions with Syn7002-RbcS nor with Syn7002-Raf1 (see 3.4). To clearly rule out that Syn7002-Raf2 does not bind to Syn7002-RbcS, a similar pull-down experiment would have to be performed with FLAG tagged Syn7002-Raf2. Moreover, Syn7002-Raf2 did not bind to pre-formed RbcL₈, while bacterial and plant Raf1 or RbcX did. The only Raf2 homolog found in cyanobacteria displays the catalytic motif of the pterin-4 α -carbinolamine dehydratase enzyme, while in plants and form IA Rubisco organisms Raf2 sequences deviate from this motif. The protein would have to be tested for its enzymatic activity to conclusively rule out that it plays a role in cyanobacterial Rubisco assembly. It has been proposed that Raf1, Raf2 and Bsd2 form transient complexes with RbcS that maintain RbcS competent for assembly with RbcL²¹⁰. Plant Raf2 contains an additional domain, not present in cyanobacterial Raf2, which may be involved in mediating these interactions. Bsd2 was the single described factor not present in the *in vitro* reconstitution

experiment (see 3.6). It was reported to be important in initial stages of Rubisco biogenesis and is present in both bundle sheath and mesophyll chloroplasts²³⁰. The *bsd2* gene is required for Rubisco accumulation in maize and Bsd2 was found to co-fractionate with polyribosomes synthesizing RbcL^{230,231}. Bsd2 contains the zinc-finger motif present in several Hsp40 proteins, a family of Hsp70 cofactors^{29,230}, suggesting a functional cooperation with Hsp70. As denatured large subunits are directly diluted to bind to chaperonin in the described *in vitro* system this factor was omitted since it is supposed to act upstream of chaperonin. Nonetheless, it cannot be ruled out that it plays a role in delivering unfolded RbcL in a partly folded and compacted state to chaperonin that might be important for proper folding. Further biochemical analysis is needed to elucidate the exact role of Raf2 and Bsd2 in Rubisco assembly and their possible functional relationship, if any, with Raf1, RbcX and other Raf proteins is yet to be analysed.

5. Outlook

Among the challenges posed by human population increase and climate change is finding ways to maximize carbon assimilation and improve crop yields. Increasing photosynthetic activity has the potential, at least in principle^{171,232}. Of the possible strategies to improve crop yield, namely, increasing the efficiency of light capture, increasing the harvest index or optimizing photosynthesis (i.e. conversion of intercepted radiation into biomass), the latter has the greater potential in the major C₃ crops¹⁷¹ since the other approaches have already been exhausted and approach their theoretical limit. Rubisco is the key limiting factor regarding photosynthetic efficiency and a prime target for improving photosynthesis.

The major obstacle towards manipulation of Rubisco is the inability to express eukaryotic Rubisco in functional form in any foreign host. There is a critical lack of knowledge regarding Rubisco assembly in higher plants. Therefore, it is essential to learn more about the chaperones involved in Rubisco folding and to characterize the intermediates leading to the holoenzyme. The biochemical and structural data presented in this thesis lead to a new model of Raf1-mediated Rubisco assembly. The next steps in characterizing important intermediates along the assembly pathway may be reached by structure-guided mutations of Raf1 and the use of heterologous components (e.g. AtRaf1.2 and Syn6301-RbcL) to produce “trapped” complexes that can be analyzed structurally.

The analysis of eukaryotic RbcX in this thesis suggests a role in Rubisco assembly similar to cyanobacterial RbcX. An important goal for the future will be the characterization of the interplay of Raf1 and RbcX in Rubisco assembly and *Synechococcus* sp. PCC7002 seems to be an excellent model system in this regard. The ultimate goal of Rubisco research will be the *in vitro* reconstitution of plant Rubisco. Consequently, detailed structural and biochemical studies regarding the role of other uncharacterized putative Rubisco assembly chaperones, e.g. Raf2 and BSD2, need to be carried out.

6. Material and Methods

6.1 Materials

6.1.1 Chemicals

Reagents used in this study had *pro analysis* quality and were delivered by **Sigma-Aldrich** (Steinheim, Germany) or **Merck** (Darmstadt, Germany) unless stated otherwise.

Amersham Biosciences (Freiburg, Germany): $\text{NaH}^{14}\text{CO}_3$

BioMol (Hamburg, Germany): IPTG

BioRad (Munich, Germany): Bradford protein assay dye reagent

Biozym (Hessisch Oldendorf, Germany): Biozym LE Agarose

Difco (Heidelberg, Germany): Bacto agar, Bacto tryptone, Bacto yeast extract

Fermentas (St. Leon-Rot, Germany): GeneRuler 1 kb DNA Ladder

Fluka (Deisenhofen, Germany): acetic acid, acetone, Bis-Tris, PEG

Invitrogen (Karlsruhe, Germany): dNTP set, protein marker for SDS-PAGE, SYBR Safe DNA gel stain, NuPAGE SDS and native PAGE gradient gels

J.M. Gabler Saliter GmbH & Co. KG (Obergünzburg, Germany): skim milk powder

MPI of Biochemistry (Martinsried, Germany), Peptide Synthesis Service: oligopeptides

Qiagen (Hilden, Germany): Ni-NTA Agarose

Roche (Basel, Switzerland): Complete protease inhibitor cocktail, DTT

Roth (Karlsruhe, Germany): ampicillin, glycine, Rotiszint

Serva (Heidelberg, Germany): Acrylamide-Bis, Coomassie blue G/R, PMSF, SDS

USB (Cleveland, USA): ammoniumsulfate, chloramphenicol, EDTA

6.1.2 Reagent and purification kits

Hampton Research (Aliso Viejo / CA, USA): Crystallization screens

Promega (Mannheim, Germany): Wizard SV Gel and PCR Clean-Up System, Wizard Plus SV Miniprep DNA Purification System, PureYield Plasmid Midiprep System, pGemTeasy vector system

Qiagen (Hilden, Germany): QIAquick PCR Purification Kit, QIAprep Spin Miniprep Kit, QIAGEN Plasmid Midi Kit

6.1.3 Enzymes, proteins and antibodies

Agrisera (Vännäs, Sweden): anti-Histone H3 antibody

Amersham Biosciences (Freiburg, Germany): porcine RNAGuard ribonuclease inhibitor

Jackson Immuno Research (Newmarket, UK): HRP-coupled secondary antibodies

JPT Peptide Technologies GmbH (Berlin, Germany): PepSpot Peptides Epitope mapping

Merck (Darmstadt, Germany): Benzonase

MPI of Biochemistry (Martinsried, Germany): antisera (produced in rabbits) against purified Syn7942-Raf1, AtRaf1.2, Syn6301-RbcL and CrRbcX-IIa, Department of Cellular Biochemistry: purified protein stocks of GroEL, GroES, Anabaena CA-RbcX

New England Biolabs (NEB, Frankfurt/Main, Germany): restriction endonucleases, T4 DNA ligase, calf intestinal phosphatase

Promega (Mannheim, Germany): Pfu DNA polymerase

Sigma-Aldrich (Steinheim, Germany): BSA, lysozyme, mouse monoclonal anti-FLAG M2 antibody, EZview Red ANTI-FLAG M2 Affinity Gel

Stratagene (La Jolla, USA): Herculase polymerase

6.1.4 Instruments

Amersham Biosciences (Freiburg, Germany): FPLC systems, prepacked chromatography columns, Protran nitrocellulose blotting membrane

Bachofer (Reutlingen, Germany): Hybridization Oven

Beckmann Coulter GmbH (Krefeld, Germany): centrifuges (J6-MI, GS-6R, Avanti 30, Avanti J-25I, Optima LE-80K ultracentrifuge), spectrophotometers (DU640, DU800), LS 6500 multi-purpose scintillation counter

Biometra (Göttingen, Germany): PCR thermocycler

Bio-Rad (Munich, Germany): Gene Pulser Xcell electroporation system, electroporation cuvettes, Mini-Protean electrophoresis chambers, Power Pac 300, Chemidoc

Branson (Connecticut, USA): Sonifier cell disruptor B15

Drummond Scientific (Broomall, USA): Pipet-aid

Eppendorf (Hamburg, Germany): centrifuges (5415D, 5417R), Thermomixer comfort, Pipettes

Fisher Scientific (Schwerte, Germany): Accumet Basic pH meter

Forma Scientific (Marietta / OH, USA): Orbital Shaker 4581

Fuji/Raytest (Straubenhardt, Germany): Fuji-LAS3000 luminescence and densitometry system

GE Healthcare (Munich, Germany): Äkta Explorer, Äkta Purifier, prepacked chromatography columns

Hampton Research (Aliso Viejo / CA, USA): Crystallization supplies and tools

Hoefer Scientific Instruments (San Francisco, USA): SEMI-PHOR blotting system

Invitrogen (Karlsruhe, Germany): XCell SureLock Mini-Cell, NuPAGE 4-12% Bis-Tris Gel 1.0 mm, 10 well, NuPAGE 3-8% Tris-Acetate Gel 1.0mm, 10 well, NuPAGE buffers

Mettler Toledo (Giessen, Germany): Balances (AG285, PB602)

Millipore (Bedford / MA, USA): Amicon Ultra centrifugal filters, Steritop vacuum filters

Misonix (Farmingdale / NY, USA): Sonicator 3000

MPI of Biochemistry (Martinsried, Germany): Western blot system (semi-dry)

MWG Biotech AG (Ebersberg, Germany): gel documentation system

New Brunswick Scientific (Nürtingen, Germany): Innova 4430 incubator

PEQLAB (Erlangen, Germany): Nanodrop 1000

Roth (Karlsruhe, Germany): ZelluTrans dialysis membrane

Scientific Industries, Inc. (Bohemia / NY, USA): Vortex-Genie 2

Whatman GmbH (Dassel, Germany): Whatman Klari-flex

WTW (Weilheim, Germany): pH meter

Wyatt Technology (Santa Barbara / CA, USA): FFF-MALS system, software ASTRA

6.1.5 Strains

Stratagene (Heidelberg, Germany): *E. coli* BL21 (DE3)

Novagen (Darmstadt, Germany): *E. coli* DH5α

6.1.6 Media

Media were prepared with deionised, double-distilled water and subsequently autoclaved.

LB medium: 10 g/l tryptone, 5 g/l yeast extract, 10 g/l NaCl,
(15 g/l agar for solid medium)

6.1.7 Plasmids and oligonucleotides

Following plasmids were constructed by S. Saschenbrecker and used in this study:

Tab.1: Plasmids used in this study from the laboratory plasmid database

| Plasmid name | genes |
|---------------------|--|
| pET11a-At-Cpn60αβ | <i>Arabidopsis thaliana</i> Chaperonin 60α (aa 47-586, at2g28000) <i>Arabidopsis thaliana</i> Chaperonin 60β (aa 56-600, at1g55490) |
| pET11a-At-Cpn60β | <i>Arabidopsis thaliana</i> Chaperonin 60β (aa 56-600, at1g55490) |
| pET11a-AtCpn20 | <i>Arabidopsis thaliana</i> Chaperonin 20 (aa 52-253, at5g20720) |
| pET11a-AtCpn10 | <i>Arabidopsis thaliana</i> Chaperonin 10 (aa 44-139, at2g44650) |
| pET11a-At-RbcL | <i>Arabidopsis thaliana</i> RbcL |
| pET11a-AnaCA-RbcX | <i>Anabeana CA</i> RbcX |
| pET11a-Syn7002-RbcS | <i>Synechococcus sp. PCC 7002</i> RbcS |
| pET11a-AtRbcS | <i>Arabidopsis thaliana</i> RbcS (aa 56-180, at1g67090) |

Plasmids generated in this study which are not described in Hauser *et al.*²⁰⁵ and Bracher and Hauser *et al.*²¹⁶ are listed in appendices (Tab.A2).

Novagen (Darmstadt, Germany): pET11a, pET30b

Metabion (Martinsried, Germany): oligonucleotides (primers)

6.2 Molecular biological methods

6.2.1 DNA analytical methods

DNA concentrations were determined by UV spectroscopy. At a wavelength of 260 nm, one absorption unit corresponds to 50 ng/μl double stranded DNA in water. The absorbance ratio 260/280 nm for pure DNA should be approximately 1.85. Deviations from this value are indicating quality deficiencies caused by impurities, such as RNA or protein²³³. Agarose gel electrophoresis was performed in 1 % TAE-agarose gels at 7-10 V/cm (TAE: 40 mM Tris-acetate, pH 8.5, 2 mM EDTA, supplemented with ethidiumbromide). Prior to loading samples on the gel, they were mixed with 5 X DNA loading dye (10mM Tris-HCl, pH 7.6, 50 mM EDTA, 0.1 % (w/v) SDS, 0.025 % (w/v) bromphenol blue, 50 % (v/v) glycerol). Authenticity of cloned constructs was confirmed by sequencing by the MPI of Biochemistry sequencing service.

6.2.2 Preparation and transformation of *E.coli* cells

Chemically competent DH5α and BL21 cells were prepared by the CaCl₂-method. Therefore, cells were grown to mid log phase (OD₆₀₀ = 0.5) at 37 °C in one litre of LB medium. After chilling the cells on ice for 10 minutes (min.), the culture was centrifuged in GSA bottles for 30 min. at 4000 rpm at 4 °C. The supernatant was removed and the cells were resuspended gently in 20 ml ice-cold 0.1 M CaCl₂, 15 % glycerol. After incubating for 20 min., the cells were aliquoted into chilled sterile eppendorf tubes, frozen in liquid nitrogen and stored at -80 °C. For transformation, 50 μl of chemically competent cells were thawed on ice and mixed with 100 ng plasmid DNA. After incubation on ice for 30 min., cells were heat shocked at 42 °C for 90 seconds (s) and put back on ice for 1 min. Next, they were diluted in 500 μl LB medium and shaken for 1 hour (h) at 37 °C. The reaction was plated out on selective LB plates and incubated overnight (o/n) at 37 °C.

6.2.3 Purification of plasmid DNA and DNA-fragments

Amplification of plasmid DNA was performed in *E. coli* DH5α, which were grown for 8-16 h in LB medium containing the appropriate antibiotics²³³. Subsequently, plasmids were purified via anion exchange chromatography using the QIAprep Spin Miniprep Kit, Wizard Plus SV Miniprep DNA Purification System, QIAGEN Plasmid Midi Kit or PureYield Plasmid Midiprep system according to the manufacturer's instructions. In order to purify or isolate DNA-fragments after agarose gel electrophoresis or enzymatic reactions, either the Wizard SV Gel and PCR Clean-Up system or the QIAquick PCR Purification Kit was used for anion exchange chromatography following the manufacturer's recommendations.

6.2.4 PCR (polymerase chain reaction)

Amplification of DNA fragments was generally performed according to protocol 1 (Table 2). Only if mutations were introduced via whole plasmid PCR²³⁴, reactions were designed with mutation-site-specific primers as listed in protocol 2 (Table 2). The applied cycling conditions are summarized in Table 3 and 4.

Tab.2: Standard protocols for PCR

| | Protocol 1 | Protocol 2 |
|--------------|---|--|
| template DNA | 7 µl plasmid (10 ng/µl) | 0.3 µl plasmid (100 ng/µl) |
| buffer | 2 µl 10 X herculase buffer | 2 µl 10 X Pfu buffer |
| dNTPs | 2 µl 2 mM dNTP | 2 µl 2 mM dNTP |
| primers | 0.5 µl 10 µM each | 0,72 µl 10 µM each |
| polymerase | 0,2 µl herculase DNA polymerase (5 U/µl) | 0,38 µl Pfu DNA polymerase (2-3 U/µl) |
| total volume | 20 µl | 20 µl |

Tab.3: Thermal cycling conditions for PCR amplification after protocol 1

| Temperature in °C | Time in minutes | Cycles |
|-------------------|-----------------|--------|
| 95 | 2 | 1 |
| 95 | 0.3 | 35 |
| 50 | 0.5 | |
| 72 | 1 min./kb | |
| 4 | ∞ | 1 |

For protocol 1, after the PCR, for some fragments (see 5.2.5), a poly A-tail was introduced by adding 2 µl 2 mM dATP after the reaction, followed by incubation at 72 °C for 30 min.

Tab.4: Thermal cycling conditions for PCR amplification after protocol 2

| Temperature in °C | Time in minutes | Cycles |
|-------------------|-----------------|--------|
| 95 | 2 | 1 |
| 95 | 0.5 | 18 |
| 55 | 1 | |
| 68 | 10 | |
| 68 | 15 | 1 |
| 4 | ∞ | 1 |

6.2.5 Restriction digest and ligation

5-20 µg DNA (PCR-product or plasmid DNA) was incubated in reactions of 20-30 µl with 20-60 units of the respective restriction enzyme/s in the absence or presence of BSA. Digests were performed in the recommended reaction buffer and at the optimal temperature for 3 h. Digested plasmid DNA was dephosphorylated with calf intestinal phosphatase according to the manufacturer's instructions prior to further use. Products of restriction digests were analyzed and isolated by agarose gel electrophoresis. For ligations, ~100 ng of digested and dephosphorylated plasmid DNA and excess of respectively digested DNA inserts were incubated with T4 DNA ligase reaction buffer and 400 U of T4 DNA ligase in a final volume of 10 µl for at least 1 h at room temperature or o/n at 16 °C. Afterwards, the ligation reactions were transformed into chemically competent *E. coli* DH5α cells. For processing of constructs resulting from whole plasmid PCR, the methylated parental DNA was digested prior to transformation. Therefore, 0,5 µl DpnI (20000 U/ml) was added to 10 µl reaction for 1 h at 37 °C²³⁴. The whole reaction was then transformed into chemically competent *E.coli* DH5α cells.

PCR fragments that were obtained after protocol 1 and finally ligated into pET11a vectors, were ligated first into pGem-T easy vector. The reaction was set up as followed: 5 µl pGem-T easy ligation buffer, 3 µl PCR fragment, 1 µl pGem-T easy vector, 1 µl T4 DNA Ligase. The mixture was incubated at RT for 1 h and finally transformed into chemically competent *E.coli* DH5α cells. After sequencing of these plasmids, the PCR fragment was cut out again and ligated into pET11a vector like described above.

6.2.6 Cloning strategies

Open reading frames for AtRbcXI, AtRbcXII, Syn7002Raf2 and AtRaf2 were amplified by PCR from genomic DNA of *Synechococcus* sp. PCC7002 (ATCC no. 27264) and *A. thaliana* cDNA,

respectively, and cloned between the SacII and SacI restriction sites of the pHue plasmid²³⁵, thus resulting in the following constructs: pHue-AtRbcXI, pHue-AtRbcXII, pHue-Syn7002Raf2 and pHueAtRaf2. The chloroplast transit peptides for AtRbcXI (1-78), AtRbcXII (1-45) and AtRaf2 (1-50) were predicted with TargetP (<http://www.cbs.dtu.dk/services/TargetP/>) or adopted from the plant proteome database (<http://ppdb.tc.cornell.edu/>). The respective primer sequences are listed in appendices (Tab.A1). For FLAG-tagged constructs, the sequence encoding the FLAG tag (MDYKDDDDKAG) was inserted upstream of the respective start codon (as described above). Point mutants were produced by whole-plasmid site-directed mutagenesis²³⁴ as described in 5.2.4 and 5.2.5. All plasmids that were generated in this study were confirmed by control digests and sequencing by the MPI of Biochemistry sequencing service. The RNAi CrrbcX-IIa silencing plasmid was constructed as described²¹⁷. The CrrbcX-IIa gene was amplified from genomic DNA of *Chlamydomonas reinhardtii* (gift from J.Nickelsen) including the transit peptide region (nucleotides 1-570) and cloned into pGem T vector easy between the EcoRI and BamHI restriction sites. The gene was cloned in a shorter version (nucleotides 1-370) behind the BamHI restriction site using EcoRI and BglII restriction sites. The final construct (nucleotides 1-570 in sense orientation followed by antisense nucleotides 1-370) was cloned between EcoRI restriction sites of plasmid NE537. The CrrbcX-IIb gene was amplified from genomic DNA and cloned between the SacII and SacI restriction sites of the pHue plasmid, resulting in the construct pHue-CrRbcX-IIb.

6.3 Protein biochemical and biophysical methods

6.3.1 Protein quantification

Protein concentrations were determined spectrophotometrically using the theoretical extinction coefficient of proteins at a wavelength of 280 nm²³⁶. Extinction coefficients were calculated by the Vector NTI program from Invitrogen. Alternatively, protein concentrations were measured colorimetrically (595 nm) by Bradford assay using the Bio-Rad protein assay reagent according to the manufacturer's recommendations^{237,238}. For proteins present as complexes in their native state (e.g. chaperones or Rubisco), molar concentrations will be expressed for oligomers, unless stated otherwise. As far as denatured substrates are concerned, concentrations will be expressed for the monomers, since this is the assumed state of substrate binding to chaperones.

6.3.2 SDS-PAGE

Proteins were analyzed under denaturing, reducing conditions by SDS-PAGE (sodiumdodecylsulfate polyacrylamide gel electrophoresis). This method allows the separation of proteins primarily according to their molecular weight and the resulting electrophoretic mobility²³⁹. Gels were cast according to Table 5. Protein samples were mixed with 5 X SDS-

loading dye (225 mM Tris-HCl, pH 6.8, 5 % SDS, 50 % (v/v) glycerol, 0.01 % (w/v) bromophenolblue, 5 % (v/v) β -mercaptoethanol) and boiled for 5 min. at 95 °C before loading onto the gels. Electrophoresis was performed in Mini-Protean electrophoresis chambers in SDS-electrophoresis buffer (12.14 mM Tris, 134.2 mM glycine, 0.1 % (w/v) SDS, 1 mM EDTA) at a constant voltage of 150 V.

Tab.5: Preparation of SDS-PAGE gels

| ingredients (amount for 4 gels) | resolution gel | | | | stacking gel |
|------------------------------------|----------------|-------------|-------------|-------------|--------------|
| | 10 % | 11 % | 12.5 % | 15 % | 4 % |
| 30 % AA/ 0.8 % BisAA | 6.8 ml | 7.33 ml | 8.4 ml | 10 ml | 1.3 ml |
| 1.5 M Tris-HCl, pH 8.8 | 5 ml | 5 ml | 5 ml | 5 ml | - |
| 0.5 M Tris-HCl, pH 6.8 | - | - | - | - | 2.5 ml |
| ddH ₂ O | 8 ml | 7.47 ml | 6.4 ml | 4.8 ml | 6.1 ml |
| 10 % (w/v) SDS | 200 μ l | 200 μ l | 200 μ l | 200 μ l | 100 μ l |
| 10 % (w/v) APS | 100 μ l | 100 μ l | 100 μ l | 100 μ l | 50 μ l |
| TEMED | 6 μ l | 6 μ l | 6 μ l | 6 μ l | 10 μ l |

6.3.3 Gradient Native PAGE

In Native PAGE the mobility of a protein depends on both its charge and its hydrodynamic size. Gels were cast according to Table 6 and samples were taken up in 5 X native loading dye (50 % (v/v) glycerol, 0.25 % (w/v) bromophenolblue, 500mM Tris pH8.6). Electrophoresis was performed in Mini-Protean electrophoresis chambers in native electrophoresis buffer (50 mM Tris, 38 mM glycin, pH not adjusted) at 4 °C, employing a constant voltage of 120 V for the first 30 min. and 150 V throughout the end of the run.

Tab.6: Preparation of native PAGE gels

| Ingredients | 4 % (3ml per gradient gel) | 13 % (3ml per gradient gel) |
|--|----------------------------|-----------------------------|
| 30 % AA/ 0.8 % BisAA | 5,33 ml | 17,4 ml |
| 4X Tris Acetate gel buffer (0.8M Tris-Acetate pH 8) | 10 ml | 10 ml |
| glycerol | 1,5 ml | 7 ml |
| dd H ₂ O | 23,17 ml | 5,6 ml |
| 10 % (w/v) APS | 15 µl | 15 µl |
| TEMED | 1,5 µl | 1,5 µl |

6.3.4 Coomassie blue staining of polyacrylamide gels

In order to detect protein amounts of $\geq 0.5 \mu\text{g}$ on SDS-PAGE, Coomassie blue staining was carried out. To fix and stain the proteins, gels were incubated in staining solution (0.16 % (w/v) Coomassie brilliant blue R-250, 40 % (v/v) ethanol, 8 % (v/v) acetic acid) followed by several washes in destaining solution (20 % (v/v) ethanol, 7 % (v/v) acetic acid) to remove background staining.

6.3.5 Western blotting and immunodetection

Immunoblotting was performed following a modified described method²⁴⁰. Proteins were transferred from gels (SDS-PAGE or native-PAGE) to a nitrocellulose membrane in a semi-dry Western blotting unit (SEMI-PHOR). The transfer was performed in Western blot buffer (20 % (v/v) methanol, 50 mM Tris, 192 mM glycine, 0.1% SDS (w/v)) at a constant current of 150 mA for 60 min. Membranes were subsequently blocked for 1 h with 10 % (w/v) milk powder in TBS buffer (20 mM Tris-HCl, pH 7.5, 137 mM NaCl) and afterwards incubated with the primary antibodies in 1 % milk TBS buffer for 1 h. Thereafter, blots were washed three times for 10 min. in TBS buffer supplemented with Tween-20 (1:1000). Then, they were incubated with the horseradish peroxidase (HRP)-conjugated secondary antibodies in 1 % milk TBS buffer for 1 h. After that, blots were washed three times for 10 min. in TBS buffer supplemented with Tween-20. For immunodetection, ECL chemiluminescence solution was freshly prepared by mixing equal amounts of ECL solution I (100 mM Tris- HCl, pH 8.5, 2.5 mM luminol (3-aminophthalhydrazide), 400 µM p-coumaric acid) and ECL solution II (100 mM Tris-HCl, pH 8.5, 5.4 mM H₂O₂). Membranes were incubated in the resulting solution and protein bands

were detected and documented with the Fuji-LAS3000 chemiluminescence and densitometry system.

6.3.6 Protein expression and purification

Protein purification was performed at 4 °C unless stated otherwise. All purification steps were monitored by SDS-PAGE and protein concentration measured by A280.

6.3.6.1 AtCpn10

E. coli BL21 (DE3) cells, transformed with pET11a-AtCpn10, were grown in LB supplemented with Ampicillin and used for expression upon induction with 1 mM IPTG (at OD₆₀₀ ~ 0.8) for 16 h at 23 °C. Harvested cells (centrifugation at 4200 rpm for 25 min.) were incubated for 45 min. in lysis buffer (50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.5 mM DTT, 0.5 mg/ml lysozyme, 10 U/ml Benzonase, complete protease inhibitor cocktail). Cells were disrupted by ultrasonication (10 x 30 s on ice, 2 min. pause). After removal of cell debris by centrifugation (40000 rpm, 40 min.), the supernatant was applied to a DEAE Sepharose fast flow column, equilibrated with 30 mM Tris-HCl, pH 7.5, 10 mM NaCl, 1mM EDTA, and eluted with a linear gradient from 10 mM to 1 M NaCl. Fractions containing AtCpn10 were desalted on a HiPrep 26/10 desalting column with 30 mM Tris-HCl pH 8.0 and applied to a MonoQ 16/10 column equilibrated with 30 mM Tris-HCl pH 8.0 and eluted with a linear gradient from 0 to 1 M NaCl. Fractions containing AtCpn10 were concentrated (MWCO 30kDa) and desalted on a HiPrep 26/10 desalting column with 30 mM Tris-HCl pH 8.0, 20 mM NaCl. The sample was then subjected to a HiPrep Sepharose fast flow 16/10 column equilibrated with 30 mM Tris-HCl pH 8.0, 20 mM NaCl and eluted with a linear gradient from 20 mM to 1 M NaCl. Fractions containing AtCpn10 were concentrated (MWCO 30 kDa) and loaded on a Superdex 200 16/10 equilibrated with 30 mM Tris-HCl pH 8.0, 20 mM NaCl. AtCpn10 containing fractions were concentrated (MWCO 30kDa), supplemented with 10 % glycerol, flash-frozen in liquid nitrogen and stored at -80 °C.

6.3.6.2 AtCpn20

Expression, cell lysis and removal of cell debris was performed as described for AtCpn10. The supernatant was subjected to a DEAE Sepharose fast flow column equilibrated with 30 mM Tris-HCl pH 7.5, 30 mM NaCl, 1 mM EDTA and eluted with a linear gradient from 10 mM to 1M NaCl. Fractions containing AtCpn20 were dialyzed against 30 mM Tris-HCl pH 7.5, 10 mM NaCl, 1mM EDTA, 1mM DTT o/n. The sample was applied to a MonoQ 16/10 column with 30 mM Tris-HCl pH 7.5, 30 mM NaCl, 1 mM EDTA, 1 mM DTT and eluted with a linear gradient from 10 mM to 1 M NaCl. The protein eluted in the flow through, was concentrated (MWCO 30 kDa) and loaded on a HiPrep Sephacryl 16/60 S-200 column equilibrated with 30 mM Tris-HCl

pH 7.5, 30 mM NaCl, 1 mM EDTA. AtCpn20 containing fractions were concentrated (MWCO 30 kDa), supplemented with 10 % glycerol, flash-frozen in liquid nitrogen and stored at -80 °C.

6.3.6.3 AtCpn60 α β ₇ / AtCpn60 β ₇

Expression, cell lysis and removal of cell debris was performed as described for AtCpn10. The supernatant was loaded on a DEAE Sepharose fast flow column equilibrated with 20 mM Tris-HCl, pH 7.5 and eluted with a linear gradient from 0 to 1 M NaCl. Fractions containing AtCpn60 α β ₇ / AtCpn60 β ₇ were desalted on a HiPrep 26/10 desalting column with 20 mM Tris-HCl pH 7.5. The sample was then subjected to a MonoQ 16/10 column equilibrated with 20 mM Tris-HCl pH 7.5 and eluted with a linear gradient from 0 to 1 M NaCl. AtCpn60 α β ₇ / AtCpn60 β ₇ containing fractions were again desalted as before and applied on a HiPrep Heparin Sepharose fast flow 16/10 column equilibrated with 20 mM Tris-HCl pH 7.5 and eluted with a linear gradient from 0 to 1 M NaCl. Fractions containing the protein of interest were concentrated (MWCO 100kDa) and loaded on a Superdex 200 26/60 column equilibrated with 20 mM Tris-HCl pH 7.5, 50 mM NaCl and 5 % glycerol. Fractions containing AtCpn60 α β ₇ / AtCpn60 β ₇ were concentrated (MWCO 100kDa), supplemented with 10 % glycerol, flash-frozen in liquid nitrogen and stored at -80 °C.

6.3.6.4 AtRaf2 / Syn7002Raf2 / FLAG-AtRbcXI / FLAG-AtRbcXII / CrRbcX-IIb

Expression, cell lysis and removal of cell debris was performed as described for AtCpn10. All proteins were expressed with an N-terminal 6XHis-Ubiquitin fusion in the pHue plasmid²³⁵. The supernatant was loaded on a Ni-NTA agarose column equilibrated with 50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 10 mM imidazole. The bound protein was washed with 50 mM Tris-HCl pH 8.0, 300 mM NaCl, 25 mM imidazole and eluted with the same buffer containing 200 mM imidazole. The 6XHis-Ubiquitin fusion tag was cleaved over night by the addition of Usp2 protease at 1:100 (mg protease: mg soluble protein) ratio in the presence of 20% (v/v) glycerol at 16 °C. The sample was desalted by a HiPrep 26/10 desalting column with 50 mM Tris-HCl pH 8.0, 50 mM NaCl and loaded on a MonoQ 16/10 equilibrated with 50 mM Tris-HCl pH 8.0, 50 mM NaCl. The protein was eluted with a linear gradient from 50 mM to 1 M NaCl. Fractions containing protein of interest were concentrated, supplemented with 10 % glycerol, flash-frozen in liquid nitrogen and stored at -80 °C.

6.3.6.5 AnaCA-RbcX

E. coli BL21 (DE3) cells, transformed with pET11a-AnaCA-rbcX, were used for expression of RbcX upon induction with 1 mM IPTG (at OD₆₀₀ ~ 0.8) for 16 h at 23 °C. Harvested cells (centrifugation at 4200 rpm for 25 min.) were incubated for 45 min. in 20 ml lysis buffer (50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.5 mM DTT, 0.5 mg/ml lysozyme, 10 U/ml Benzonase, complete protease inhibitor cocktail). Cells were disrupted by ultrasonication (10 x 30 s on ice,

2 min. pause). After removal of cell debris by centrifugation (40000 rpm, 40 min.), the supernatant was applied to a SourceQ column, equilibrated with 50 mM Tris-HCl, pH 8.0, 0.5 mM DTT, and eluted with a linear gradient from 0 to 1 M NaCl. Fractions containing RbcX were dialyzed against 20 mM imidazole, pH 6.4, 20 mM NaCl, applied to an equilibrated MonoQ column, and eluted with a linear salt gradient up to 0.7 M NaCl. Fractions containing RbcX were dialyzed against 20 mM Tris-HCl, pH 7.5, 50 mM NaCl, 10 mM MgCl₂. RbcX was then concentrated (MWCO 10 kDa) and subjected to Superdex 200 gel filtration chromatography. Fractions containing RbcX were concentrated (MWCO 10 kDa), supplemented with 10 % glycerol, flash-frozen in liquid nitrogen and stored at -80 °C.

6.3.6.6 Prokaryotic and Eukaryotic RbcL

Syn6301-RbcL and At-RbcL were purified in denatured form from inclusion bodies after expression in *E. coli* BL21 (DE3) cells transformed with plasmid pET11a encoding the respective *rbcL* gene. Cells were grown in LB amp at 37 °C to an OD600 of 0.8 and protein expression was induced by adding 1mM IPTG for 5 hr at 30 °C. Harvested cells (30 min., 4200 rpm) were resuspended in 40 mM Tris-HCl, pH 8.0, 0.25 M sucrose, 1% (v/v) Triton X-100, supplemented with Complete protease inhibitor cocktail and 0.5 mg/ml lysozyme and incubated on ice for 30 min., whereupon 0.5 M EDTA was added. Cells were disrupted by ultrasonication (10 x 30 s on ice, 2 min. pause) and inclusion bodies pelleted by centrifugation (1 hr, 22,000 g). Inclusion bodies were transferred to a homogenizer and resuspended in 40 mM Tris-HCl, pH 8.0, 0.25 M sucrose, 10 mM EDTA, 1% Triton X-100, 2 M Urea. Washed inclusion bodies were centrifuged (45 min, 40000 rpm) and again homogenized with 40 mM Tris-HCl, pH 8.0, 0.25 M sucrose, 10 mM EDTA followed by centrifugation (45 min., 40000rpm). In the final step, the purified denatured protein was resuspended in 40 mM Tris-HCl, pH 8.0, 6 M GdnHCl, 1 mM EDTA, 5 mM DTT. The protein was flash frozen in liquid nitrogen and stored at -80 °C.

6.3.6.7 Prokaryotic and Eukaryotic RbcS

Expression, disruption, and purification of RbcS from *E. coli* BL21 (DE3) cells was performed essentially as described for the purification of RbcL from inclusion bodies^{241,242}. However, after the final resuspension, denatured purified protein was diluted in 40 mM Tris-HCl, pH 8.0, 6 M GdnHCl, 1 mM EDTA, 5 mM DTT to a final protein concentration of 0.5 mg/ml and dialyzed against 50 mM Tris-HCl, pH 8.0, 50 mM NaCl, 10 mM MgCl₂, where the protein could spontaneously refold into its native state. Aggregated protein was removed by ultracentrifugation (30 min., 40,000 rpm), refolded purified protein was further concentrated (MWCO 10kDa), supplemented with 10 % glycerol, flash-frozen in liquid nitrogen and stored at -80 °C.

6.3.7 Co-immunoprecipitation

All centrifugation steps with FLAG beads were performed for 30 s at 8,200 g and 4 °C, beads were kept at 4 °C, and all incubation steps were performed with gentle rotation. Before use, gel beads were washed twice with 1XTBS-T (50 mM Tris-HCl, pH 8.0, 137 mM NaCl, 2.7 mM KCl, 0.1% Tween-20) buffer and vortexing. Proteins that were tested for interaction with each other (2.5 μ M RbcS, 1 μ M Raf1 and 1 μ M Raf2 for respective experiments) were incubated with 40 μ l anti-FLAG M2 Affinity Gel beads for 3 hr at 4 °C under rotation in 300 μ l 20 mM MOPS-KOH, pH 7.5, 100 mM KCl, 5 mM Mg-acetate, 5 mM DTT in the absence or presence of BSA (0.8 mg/ml). The beads were washed 3 times 10 min. with 1XTBS-T and bound protein was eluted by incubating the beads for 5 min at 95 °C in non-reducing SDS-sample buffer (no β -mercapthoethanol) and analyzed on 12.5 % SDS-PAGE gels, followed by coomassie staining.

6.3.8 Peptide binding screen

Analysis of Raf1 binding to RbcL and RbcX binding to the C-terminal peptide region of RbcL was performed using a peptide array membrane prepared by JPT Peptide Technologies GmbH; the epitope map is prepared by SPOT-synthesis. The tri- or dodecapeptides, with 10 amino acid overlap covering the sequence of RbcL or the last 40 amino acids in various RbcL species, respectively, are covalently bound to Whatman 50 cellulose support by the C-terminus end and have an acetylated N-terminus that is not bound. Each spot contains approximately 5 nmol of peptide. Before protein was bound to the membrane, and after each membrane regeneration, the membrane was tested for background binding of primary or secondary antibody to the spots or if residual protein was still bound after regeneration. Blank detection was begun by washing the membrane for 5 min. with 100% methanol. The methanol wash is performed to avoid hydrophobic peptide precipitation in the following TBS washing steps. The membrane was then washed 3 times 5 min. with 1xTBS-T and blocked in 10X milk in TBS for 3h. The membrane was again washed 3 times 5 min. with 1xTBS-T and incubated for 45 min. with 1:2,000-5,000 dilutions of primary antibody in 1X milk in TBS. Excess antibody was removed by washing the membrane 3 times 5 min. with 1xTBS-T. Then, the membrane was incubated with 1:5,000 dilutions of HRP-conjugated secondary antibody in 1X milk with TBS for 45 min. Excess antibody was removed by washing 3 times 5 min. with 1xTBS-T and chemiluminescent detection performed as described in 5.3.5. Incubation of the membrane with respective proteins was performed as described above, by overnight incubation of 10 μ g/ml protein in 1X milk in TBS. After protein incubation and ECL detection, the protein bound to the membrane was removed by regenerating the membrane. The membrane was incubated at 50 °C 4 times 30 min. in 62.5 mM Tris-HCl pH 6.7, 2% (w/v) SDS, 100 mM β -mercaptoethanol. The membrane was washed 3 times for 20 min. with 10xPBS, followed by washing 3 times 20 min. with 1xTBS-T. The membrane could then be blank detected to determine the extent of

protein removal. Regeneration of the membrane was repeated until all protein was removed or only minimal amounts of protein remained bound.

6.3.9 Transformation and selection of *Chlamydomonas reinhardtii*

C. reinhardtii cells were grown in Trisacetate-phosphate (TAP) medium²⁴³. The strain UVM4²⁴⁴ was transformed by the glass beads procedure²⁴⁵ and allowed to recover for 2 days, to permit induction of RNAi, before plating under selective conditions. Transformants were selected on TAP medium containing 5 g/ml paromomycin. Plates were incubated under dim lights (approximately 50 mol / m²s (photosynthetically active radiation) covered with one layer of paper towels²⁴⁶. Isolated transgenic strains were kept under constant selective pressure to circumvent any silencing of integrated IR transgenes.

6.3.10 *Chlamydomonas* total RNA isolation

All steps were performed in an RNase free environment at 4 °C. Total cell RNA was purified with TRI Reagent (Sigma Aldrich), according to the manufacturer's instructions. Reverse transcriptase reactions were performed using 500 ng of isolated RNA with the following PCR protocol : 5 min. at 25 °C, 30 min. at 42 °C, 5 min. at 85 °C and stored at 4 °C. Synthesized cDNA was then used as a template in standard PCR reactions (28 cycles, 5.2.4). 15 µl aliquots of each RT-PCR reaction were resolved on 1.5 % agarose gels and visualized by sybre safe staining (primers see appendix Tab.A1).

6.3.11 *Chlamydomonas* total protein analysis

100 ml of transformant cells were grown to mid log phase. Pelleted cells (15 min., 20000 g) were resuspended in 2 ml 50 mM bicine pH 8.0, 100 mM NaHCO₃, 10 mM MgCl₂. After incubating pellets on ice for 30 min., cells were lysed by ultrasonication (3 times 30s, pause on ice inbetween) and cell debris was removed by centrifugation (30 min., 13200 rpm). Potein concentration was measured by Bradford. Samples were taken for SDS-PAGE analysis by mixing with SDS loading dye at ech step.

7. References

1. Anfinsen, C.B., Redfield, R.R., Choate, W.L., Page, J., & Carroll, W.R. Studies on the gross structure, cross-linkages, and terminal sequences in ribonuclease. *J Biol Chem* **207**, 201–210 (1954).
2. Anfinsen, C.B. & Haber, E. Studies on the reduction and re-formation of protein disulfide bonds. *J Biol Chem* **236**, 1361–1363 (1961).
3. Haber, E. & Anfinsen, C.B. Side-chain interactions governing pairing of half-cystine residues in ribonuclease. *J. Biol. Chem.* **237**, 1839–& (1962).
4. Anfinsen, C.B. Principles that govern folding of protein chains. *Science* **181**, 223–230 (1973).
5. Mayer, M.P. Gymnastics of molecular chaperones. *Mol. Cell* **39**, 321–331 (2010).
6. Levinthal, C. Are there pathways for protein folding. *J. Chim. Phys. Physico-Chimie Biol.* **65**, 44–45 (1968).
7. Kim, Y.E., Hipp, M.S., Bracher, A., Hayer-Hartl, M., & Hartl, F.U. Molecular chaperone functions in protein folding and proteostasis. *Annu. Rev. Biochem. Vol 82* **82**, 323–355 (2013).
8. Dobson, C.M. & Karplus, M. The fundamentals of protein folding: bringing together theory and experiment. *Curr Opin Struct Biol* **9**, 92–101 (1999).
9. Ellis, R.J. & Minton, A.P. Cell biology: join the crowd. *Nature* **425**, 27–28 (2003).
10. Ellis, R.J. Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Curr Opin Struct Biol* **11**, 114–119 (2001).
11. Ellis, R.J. & Hemmingsen, S.M. Molecular chaperones: proteins essential for the biogenesis of some macromolecular structures. *Trends Biochem Sci* **14**, 339–342 (1989).
12. Young, J.C., Agashe, V.R., Siegers, K., & Hartl, F.U. Pathways of chaperone-mediated protein folding in the cytosol. *Nat. Rev. Mol. Cell Biol.* **5**, 781–791 (2004).
13. Hartl, F.U. & Hayer-Hartl, M. Protein folding - molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295**, 1852–1858 (2002).
14. Ellis, R.J., Henderson, B. & Pockley, A.G. Molecular chaperones: the orthodox view. In *Molecular Chaperones and Cell Signalling* (ed. Henderson, B., & Pockley, A.G.) 3–21 (Cambridge University Press, 2005).
15. Frydman, J. Folding of newly translated proteins in vivo: the role of molecular chaperones. *Annu. Rev. Biochem. Vol 82* **70**, 603–647 (2001).
16. Brandt, F., Carlson, L.A., Hartl, F.U., Baumeister, W., & Grunewald, K. The three-dimensional organization of polyribosomes in intact human cells. *Mol. Cell* **39**, 560–569 (2010).
17. Brandt, F. *et al.* The native 3D organization of bacterial polysomes. *Cell* **136**, 261–271 (2009).
18. Langer, T. *et al.* Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. *Nature* **356**, 683–689 (1992).

19. Frydman, J., Nimmesgern, E., Ohtsuka, K., & Hartl, F.U. Folding of nascent polypeptide chains in a high molecular mass assembly with molecular chaperones. *Nature* **370**, 111–117 (1994).
20. Rutherford, S.L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342 (1998).
21. Tokuriki, N. & Tawfik, D.S. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* **459**, 668–673 (2009).
22. Bukau, B. & Horwich, A.L. The hsp70 and hsp60 chaperone machines. *Cell* **92**, 351–366 (1998).
23. Zuiderweg, E.R. *et al.* Allostery in the hsp70 chaperone proteins. *Top Curr Chem* **328**, 99–153 (2013).
24. Hartl, F.U., Bracher, A., & Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature* **475**, 324–332 (2011).
25. Kityk, R., Kopp, J., Sinning, I., & Mayer, M.P. Structure and dynamics of the ATP-bound open conformation of hsp70 chaperones. *Mol. Cell* **48**, 863–874 (2012).
26. Zhuravleva, A. & Gierasch, L.M. Allosteric signal transmission in the nucleotide-binding domain of 70-kda heat shock protein (hsp70) molecular chaperones. *Proc Natl Acad Sci U S A* **108**, 6987–6992 (2011).
27. Bertelsen, E.B., Chang, L., Gestwicki, J.E., & Zuiderweg, E.R. Solution conformation of wild-type E. coli hsp70 (DnaK) chaperone complexed with ADP and substrate. *Proc Natl Acad Sci U S A* **106**, 8471–8476 (2009).
28. Mapa, K. *et al.* The conformational dynamics of the mitochondrial hsp70 chaperone. *Mol. Cell* **38**, 89–100 (2010).
29. Kampinga, H.H. & Craig, E.A. The hsp70 chaperone machinery: J proteins as drivers of functional specificity. *Nat Rev Mol Cell Biol* **11**, 579–592 (2010).
30. Hartl, F.U. & Hayer-Hartl, M. Converging concepts of protein folding in vitro and in vivo. *Nat. Struct. Mol. Biol.* **16**, 574–581 (2009).
31. Jiang, J. *et al.* Structural basis of j cochaperone binding and regulation of hsp70. *Mol. Cell* **28**, 422–433 (2007).
32. Ahmad, A. *et al.* Heat shock protein 70 kda chaperone/DnaJ cochaperone complex employs an unusual dynamic interface. *Proc Natl Acad Sci U S A* **108**, 18966–18971 (2011).
33. Svärd, M., Biterova, E.I., Bourhis, J.M., & Guy, J.E. The crystal structure of the human co-chaperone p58 (IPK). *PLoS One* **6**, e22337 (2011).
34. Hagiwara, M. *et al.* Structural basis of an ERAD pathway mediated by the ER-resident protein disulfide reductase ERdj5. *Mol. Cell* **41**, 432–444 (2011).
35. Klumpp, M., Baumeister, W., & Essen, L.O. Structure of the substrate binding domain of the thermosome, an archaeal group II chaperonin. *Cell* **91**, 263–270 (1997).
36. Ditzel, L. *et al.* Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT. *Cell* **93**, 125–138 (1998).
37. Leitner, A. *et al.* The molecular architecture of the eukaryotic chaperonin TRiC/CCT. *Structure* **20**, 814–825 (2012).

References

38. Mayhew, M. *et al.* Protein folding in the central cavity of the GroEL-GroES chaperonin complex. *Nature* **379**, 420–426 (1996).
39. Walter, S. Structure and function of the GroE chaperone. *Cell Mol Life Sci* **59**, 1589–1597 (2002).
40. Hartl, F.U. Molecular chaperones in cellular protein folding. *Nature* **381**, 571–579 (1996).
41. Sigler, P.B. *et al.* Structure and function in GroEL-mediated protein folding. *Annu. Rev. Biochem. Vol 82* **67**, 581–608 (1998).
42. Xu, Z., Horwich, A.L., & Sigler, P.B. The crystal structure of the asymmetric GroEL-GroES-(ADP)₇ chaperonin complex. *Nature* **388**, 741–750 (1997).
43. Richardson, A., Landry, S.J., & Georgopoulos, C. The ins and outs of a molecular chaperone machine. *Trends Biochem Sci* **23**, 138–143 (1998).
44. Saibil, H.R. *et al.* ATP induces large quaternary rearrangements in a cage-like chaperonin structure. *Curr Biol* **3**, 265–273 (1993).
45. Ewalt, K.L., Hendrick, J.P., Houry, W.A., & Hartl, F.U. In vivo observation of polypeptide flux through the bacterial chaperonin system. *Cell* **90**, 491–500 (1997).
46. Walter, S. & Buchner, J. Molecular chaperones-cellular machines for protein folding. *Angew Chem Int Ed Engl* **41**, 1098–1113 (2002).
47. Houry, W.A., Frishman, D., Eckerskorn, C., Lottspeich, F., & Hartl, F.U. Identification of in vivo substrates of the chaperonin GroEL. *Nature* **402**, 147–154 (1999).
48. Kerner, M.J. *et al.* Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell* **122**, 209–220 (2005).
49. Brinker, A. *et al.* Dual function of protein confinement in chaperonin-assisted protein folding. *Cell* **107**, 223–233 (2001).
50. Thirumalai, D. & Lorimer, G.H. Chaperonin-mediated protein folding. *Annu Rev Biophys Biomol Struct* **30**, 245–269 (2001).
51. Tang, Y.C. *et al.* Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein. *Cell* **125**, 903–914 (2006).
52. Lin, Z., Madan, D., & Rye, H.S. GroEL stimulates protein folding through forced unfolding. *Nat. Struct. Mol. Biol.* **15**, 303–311 (2008).
53. Chakraborty, K. *et al.* Chaperonin-catalyzed rescue of kinetically trapped states in protein folding. *Cell* **142**, 112–122 (2010).
54. Baumketner, A., Jewett, A., & Shea, J.E. Effects of confinement in chaperonin assisted protein folding: rate enhancement by decreasing the roughness of the folding energy landscape. *J Mol Biol* **332**, 701–713 (2003).
55. Hayer-Hartl, M. & Minton, A.P. A simple semiempirical model for the effect of molecular confinement upon the rate of protein folding. *Biochemistry* **45**, 13356–13360 (2006).
56. England, J., Lucent, D., & Pande, V. Rattling the cage: computational models of chaperonin-mediated protein folding. *Curr Opin Struct Biol* **18**, 163–169 (2008).

57. Tian, J. & Garcia, A.E. Simulation studies of protein folding/unfolding equilibrium under polar and nonpolar confinement. *J Am Chem Soc* **133**, 15157–15164 (2011).
58. Gupta, A.J., Haldar, S., Milicic, G., Hartl, F.U., & Hayer-Hartl, M. Active cage mechanism of chaperonin-assisted protein folding demonstrated at single-molecule level. *J Mol Biol* **426**, 2739–2754 (2014).
59. Apetri, A.C. & Horwich, A.L. Chaperonin chamber accelerates protein folding through passive action of preventing aggregation. *Proc Natl Acad Sci U S A* **105**, 17351–17355 (2008).
60. Tang, Y.C., Chang, H.C., Chakraborty, K., Hartl, F.U., & Hayer-Hartl, M. Essential role of the chaperonin folding compartment in vivo. *EMBO J.* **27**, 1458–1468 (2008).
61. Wang, J.D., Herman, C., Tipton, K.A., Gross, C.A., & Weissman, J.S. Directed evolution of substrate-optimized GroEL/S chaperonins. *Cell* **111**, 1027–1039 (2002).
62. Laskey, R.A., Honda, B.M., Mills, A.D., & Finch, J.T. Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. *Nature* **275**, 416–420 (1978).
63. Luger, K., Maeder, A.W., Richmond, R.K., Sargent, D.F., & Richmond, T.J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–259 (1997).
64. Luger, K. Dynamic nucleosomes. *Chromosome Res.* **14**, 5-16 (2006).
65. Haushalter, K.A., & Kadonaga, J.T. Chromatin assembly by DNA-translocating motors. *Nat. Rev. Mol. Cell Biol.* **4**, 613–620 (2003).
66. Hochstrasser, M. Ubiquitin-dependent protein degradation. *Annu. Rev. Genet.* **30**, 405-439 (1996).
67. Goldberg, A.L. Functions of the proteasome: from protein degradation and immune surveillance to cancer therapy. *Biochem. Soc. Trans.* **35**, 12–17 (2007).
68. Schmidt, M., Hanna, J., Elsasser, S., & Finley, D. Proteasome-associated proteins: regulation of a proteolytic machine. *Biol. Chem.* **386**, 725–737 (2005).
69. Hirano, Y. *et al.* A heterodimeric complex that promotes the assembly of mammalian 20S proteasomes. *Nature* **437**, 1381–1385 (2005).
70. Chen, P., & Hochstrasser, M. Autocatalytic subunit processing couples active site formation in the 20S proteasome to completion of assembly. *Cell* **86**, 961–972 (1996).
71. Kusmierczyk, A.R., Kunjappu, M.J., Funakoshi, M., & Hochstrasser, M. A multimeric assembly factor controls the formation of alternative 20S proteasomes. *Nat. Struct. Mol. Biol.* **15**, 237–244 (2008).
72. Funakoshi, M., Tomko, R.J., Kobayashi, H., & Hochstrasser, M. Multiple assembly chaperones govern biogenesis of the proteasome regulatory particle base. *Cell* **137**, 887-899 (2009).
73. Ansel, K.M. *et al.* Mouse Eri1 interacts with the ribosome and catalyzes 5.8S rRNA processing. *Nat. Struct. Mol. Biol.* **15**, 523–530 (2008).
74. Gabel, H.W. & Ruvkun, G. The exonuclease ERI-1 has a conserved dual role in 5.8S rRNA processing and RNAi. *Nat. Struct. Mol. Biol.* **15**, 531–533 (2008).
75. Thomson, E. & Tollervey, D. The final step in 5.8S rRNA processing is cytoplasmic in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **30**, 976-984 (2009).

References

76. Sauer, F.G., Remaut, H., Hultgren, S.J., & Waksman, G. Fiber assembly by the chaperone-usher pathway. *Biochim. Biophys. Acta* **1694**, 259–267 (2004).
77. Barnhart, M.M. *et al.* PapD-like chaperones provide the missing information for folding of pilin proteins. *Proc. Natl Acad. Sci.* **97**, 7709–7714 (2000).
78. Hultgren, S.J. *et al.* The PapG adhesin of uropathogenic *Escherichia coli* contains separate regions for receptor binding and for the incorporation into the pilus. *Proc. Natl Acad. Sci.* **86**, 4357–4361 (1989).
79. Vetsch, M. *et al.* Pilus chaperones represent a new type of protein-folding catalyst. *Nature* **431**, 329–333 (2004).
80. Holmgren, A., & Branden, C.I. Crystal structure of chaperone protein PapD reveals an immunoglobulin fold. *Nature* **342**, 248–251 (1989).
81. Sauer, F.G. *et al.* Structural basis of chaperone function and pilus biogenesis. *Science* **285**, 1058–1061 (1999).
82. Soll, J. & Schleiff, E. Protein import into chloroplasts. *Nat Rev Mol Cell Biol* **5**, 198–208 (2004).
83. Whatley, F.R., Tagawa, K., & Arnon, D.I. Separation of the light and dark reactions in electron transfer during photosynthesis. *Proc Natl Acad Sci U S A* **49**, 266–270 (1963).
84. Hill, R. & Bendall, F. Function of the 2 cytochrome components in chloroplasts - working hypothesis. *Nature* **186**, 136–137 (1960).
85. Duysens, L.N., Kamp, B.M., & Ames, J. Two photochemical systems in photosynthesis. *Nature* **190**, 510–& (1961).
86. Mitchell, P. Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature* **191**, 144–148 (1961).
87. McCarty, R.E. & Racker, E. Effect of a coupling factor and its antiserum on photophosphorylation and hydrogen ion transport. *Brookhaven Symp. Biol.* **19**, 202-214 (1966).
88. Jagendorf, A.T. Acid-base transitions and phosphorylation by chloroplasts. *Fed. Proc.* **26**, 1361-1369 (1967).
89. Cramer, W.A. & Butler, W.L. Light-induced absorbance changes of 2 cytochrome b components in electron-transport system of spinach chloroplasts. *Biochim. Biophys. Acta* **143**, 332–& (1967).
90. Kaftan, D., Brumfeld, V., Nevo, R., Scherz, A., & Reich, Z. From chloroplasts to photosystems: in situ scanning force microscopy on intact thylakoid membranes. *EMBO J.* **21**, 6146–6153 (2002).
91. Junge, W. ATP synthase and other motor proteins. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 4735–4737 (1999).
92. Herrmann, R.G. Biogenesis and evolution of photosynthetic (thylakoid) membranes. *Biosci. Rep.* **19**, 355–365 (1999).
93. Joliot, P. & Joliot, A. Cyclic electron transfer in plant leaf. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 10209–10214 (2002).
94. Eberhard, S., Finazzi, G., & Wollman, F.A. The dynamics of photosynthesis. *Annu. Rev. Genet.* **42**, 463–515 (2008).

95. Lichtenthaler, H.K. The 1-deoxy-d-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**, 47–65 (1999).
96. Bowes, G., Ogren, W.L., & Hageman, R.H. Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase. *Biochem. Biophys. Res. Commun.* **45**, 716–& (1971).
97. Lorimer, G.H., Andrews, T.J., & Tolbert, N.E. Ribulose diphosphate oxygenase. 2.further proof of reaction-products and mechanism of action. *Biochemistry* **12**, 18–23 (1973).
98. Monteith, J.L. Climate and efficiency of crop production in britain. *Philos. Trans. R. Soc. London Ser. B-Biological Sci.* **281**, 277–294 (1977).
99. Gutteridge, S., Parry, M.A.J., Schmidt, C.N.G., & Feeney, J. An investigation of ribulosebiphosphate carboxylase activity by high-resolution H-1-NMR. *Febs Lett.* **170**, 355–359 (1984).
100. Osmond, B., Badger, M., Maxwell, K., Bjorkman, O., & Leegood, R. Too many photos: photorespiration, photoinhibition and photooxidation. *Trends Plant Sci.* **2**, 119–121 (1997).
101. Wu, J., Neimanis, S., & Heber, U. Photorespiration is more effective than the Mehler reaction in protecting the photosynthetic apparatus against photoinhibition. *Bot. Acta* **104**, 283–291 (1991).
102. Osmond, C.B. & Grace, S.C. Perspectives on photoinhibition and photorespiration in the field - quintessential inefficiencies of the light and dark reactions of photosynthesis. *J. Exp. Bot.* **46**, 1351–1362 (1995).
103. Kozaki, A. & Takeba, G. Photorespiration protects C3 plants from photooxidation. *Nature* **384**, 557–560 (1996).
104. Madore, M. & Grodzinski, B. Effect of oxygen concentration on ¹⁴C-photoassimilate transport from leaves of *Salvia splendens* L. *Plant Physiol.* **76**, 782–786 (1984).
105. Noctor, G. & Foyer, C.H. Ascorbate and glutathione: keeping active oxygen under control. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 249–279 (1998).
106. Noctor, G., Arisi, A.C.M., Jouanin, L., & Foyer, C.H. Photorespiratory glycine enhances glutathione accumulation in both the chloroplastic and cytosolic compartments. *J. Exp. Bot.* **50**, 1157–1167 (1999).
107. Noctor, G. *et al.* Glutathione: biosynthesis, metabolism and relationship to stress tolerance explored in transformed plants. *J. Exp. Bot.* **49**, 623–647 (1998).
108. Wingler, A., Lea, P.J., Quick, W.P., & Leegood, R.C. Photorespiration: metabolic pathways and their role in stress protection. *Philos. Trans. R. Soc. B-Biological Sci.* **355**, 1517–1529 (2000).
109. Price, G.D., Badger, M.R., Woodger, F.J., & Long, B.M. Advances in understanding the cyanobacterial CO₂-concentrating-mechanism (ccm): functional components, C_i transporters, diversity, genetic regulation and prospects for engineering into plants. *J. Exp. Bot.* **59**, 1441–1461 (2008).
110. Badger, M.R., Hanson, D., & Price, G.D. Evolution and diversity of CO₂ concentrating mechanisms in cyanobacteria. *Funct. Plant Biol.* **29**, 161–173 (2002).
111. Badger, M.R. & Price, G.D. CO₂ concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. *J. Exp. Bot.* **54**, 609–622 (2003).
112. Heinhorst, S. *et al.* Characterization of the carboxysomal carbonic anhydrase CsoSCA from *Halothiobacillus neapolitanus*. *J Bacteriol* **188**, 8087–8094 (2006).

References

113. Dou, Z. *et al.* CO₂ fixation kinetics of *Halothiobacillus neapolitanus* mutant carboxysomes lacking carbonic anhydrase suggest the shell acts as a diffusional barrier for CO₂. *J Biol Chem* **283**, 10377–10384 (2008).
114. Yeates, T.O., Kerfeld, C.A., Heinhorst, S., Cannon, G.C., & Shively, J.M. Protein-based organelles in bacteria: carboxysomes and related microcompartments. *Nat Rev Microbiol* **6**, 681–691 (2008).
115. Morita, E., Kuroiwa, H., Kuroiwa, T., & Nozaki, H. High localization of ribulose-1,5-bisphosphate carboxylase/oxygenase in the pyrenoids of *Chlamydomonas reinhardtii* (chlorophyta), as revealed by cryofixation and immunogold electron microscopy. *J. Phycol.* **33**, 68–72 (1997).
116. Edwards, G.E., Ku, M.S.B. & Monson, R.K. C₄ photosynthesis and its regulation. In *Photosynthetic Mechanisms and the Environment* (Elsevier Science Ltd, 1985).
117. Griffiths, H. Carbon-dioxide concentrating mechanisms and the evolution of CAM in vascular epiphytes. *Vasc. Plants as Epiphytes* **76**, 42–86 (1989).
118. Andrews, T.J. & Lorimer, G.H. Rubisco: structure, mechanisms, and prospects for improvement. In *The Biochemistry of Plants* (ed. Hatch, M.D. & Boardman, N. K.) **10**, 131–218 (Academic Press, 1987).
119. Ellis, R.J. Most abundant protein in the world. *Trends Biochem. Sci.* **4**, 241–244 (1979).
120. Evans, M.C.W., Buchanan, B.B., & Arnon, D.I. A new ferredoxin-dependent carbon reduction cycle in a photosynthetic bacterium. *Proc. Natl. Acad. Sci. U. S. A.* **55**, 928–& (1966).
121. Takabe, T. & Akazawa, T. A comparative study on the effect of O₂ on photosynthetic carbon metabolism by *Chlorobium thiosulfatophilum* and *Chromatium vinosum*. *Plant Cell Physiol.* **18**, 753–765 (1977).
122. McFadden, B.A. Autotrophic CO₂ assimilation and evolution of ribulose diphosphate carboxylase. *Bacteriol. Rev.* **37**, 289–319 (1973).
123. Kawashima, N. & Wildman, S.G. Studies on fraction-I protein. 1.Effect of crystallization of fraction-I protein from tobacco leaves on ribulose diphosphate carboxylase activity. *Biochim. Biophys. Acta* **229**, 240–& (1971).
124. Schneider, G., Lindqvist, Y., Branden, C.I., & Lorimer, G. 3-dimensional structure of ribulose-1,5-bisphosphate carboxylase-oxygenase from *Rhodospirillum-rubrum* at 2.9 Å resolution. *EMBO J.* **5**, 3409–3415 (1986).
125. Somerville, C.R. & Somerville, S.C. Cloning and expression of the *Rhodospirillum-rubrum* ribulosebiphosphate carboxylase gene in *Escherichia-coli*. *Mol. Gen. Genet.* **193**, 214–219 (1984).
126. Andersson, I. *et al.* Crystal-structure of the active-site of ribulose-bisphosphate carboxylase. *Nature* **337**, 229–234 (1989).
127. Knight, S., Andersson, I., & Branden, C.I. Crystallographic analysis of ribulose 1,5-bisphosphate carboxylase from spinach at 2.4 Å resolution - subunit interactions and active-site. *J. Mol. Biol.* **215**, 113–160 (1990).
128. Curmi, P.M.G., Cascio, D., Sweet, R.M., Eisenberg, D., & Schreuder, H. Crystal-structure of the unactivated form of ribulose-1,5-bisphosphate carboxylase oxygenase from tobacco refined at 2.0-Å resolution. *J. Biol. Chem.* **267**, 16980–16989 (1992).

129. Schreuder, H.A. *et al.* Crystal-structure of activated tobacco rubisco complexed with the reaction-intermediate analog 2-carboxy-arabinitol 1,5-bisphosphate. *Protein Sci.* **2**, 1136–1146 (1993).
130. Newman, J., Branden, C.I., & Jones, T.A. Structure determination and refinement of ribulose 1,5-bisphosphate carboxylase/oxygenase from *Synechococcus* PCC6301. *Acta Crystallogr. Sect. D-Biological Crystallogr.* **49**, 548–560 (1993).
131. Newman, J. & Gutteridge, S. The x-ray structure of *Synechococcus* ribulose-bisphosphate carboxylase oxygenase-activated quaternary complex at 2.2-Angstrom resolution. *J. Biol. Chem.* **268**, 25876–25886 (1993).
132. Delwiche, C.F. & Palmer, J.D. Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids. *Mol. Biol. Evol.* **13**, 873–882 (1996).
133. Tabita, F.R. Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: a different perspective. *Photosynth. Res.* **60**, 1–28 (1999).
134. Kitano, K. *et al.* Crystal structure of a novel-type archaeal rubisco with pentagonal symmetry. *Structure* **9**, 473–481 (2001).
135. Tabita, F.R., Hanson, T.E., Satagopan, S., Witte, B.H., & Kreel, N.E. Phylogenetic and evolutionary relationships of rubisco and the rubisco-like proteins and the functional lessons provided by diverse molecular forms. *Philos. Trans. R. Soc. B-Biological Sci.* **363**, 2629–2640 (2008).
136. Li, H.Y., Sawaya, M.R., Tabita, F.R., & Eisenberg, D. Crystal structure of a rubisco-like protein from the green sulfur bacterium *Chlorobium tepidum*. *Structure* **13**, 779–789 (2005).
137. Tabita, F.R. *et al.* Function, structure, and evolution of the rubisco-like proteins and their rubisco homologs. *Microbiol. Mol. Biol. Rev.* **71**, 576–599 (2007).
138. Imker, H.J., Fedorov, A.A., Fedorov, E. V, Almo, S.C., & Gerlt, J.A. Mechanistic diversity in the rubisco superfamily: the 'enolase' in the methionine salvage pathway in *Geobacillus kaustophilus*. *Biochemistry* **46**, 4077–4089 (2007).
139. Andersson, I. & Taylor, T.C. Structural framework for catalysis and regulation in ribulose-1,5-bisphosphate carboxylase/oxygenase. *Arch. Biochem. Biophys.* **414**, 130–140 (2003).
140. Schneider, G. *et al.* Comparison of the crystal-structures of L2 and L8S8 rubisco suggests a functional-role for the small subunit. *EMBO J.* **9**, 2045–2050 (1990).
141. Spreitzer, R.J. Role of the small subunit in ribulose-1,5-bisphosphate carboxylase/oxygenase. *Arch. Biochem. Biophys.* **414**, 141–149 (2003).
142. Taylor, T.C., Backlund, A., Bjorhall, K., Spreitzer, R.J., & Andersson, I. First crystal structure of rubisco from a green alga, *Chlamydomonas reinhardtii*. *J. Biol. Chem.* **276**, 48159–48164 (2001).
143. Andersson, I. & Backlund, A. Structure and function of rubisco. *Plant Physiol. Biochem.* **46**, 275–291 (2008).
144. Jordan, D.B. & Ogren, W.L. Species variation in the specificity of ribulose-biphosphate carboxylase-oxygenase. *Nature* **291**, 513–515 (1981).
145. Badger, M.R. & Lorimer, G.H. Activation of ribulose-1,5-bisphosphate oxygenase - role of Mg²⁺, CO₂, and pH. *Arch. Biochem. Biophys.* **175**, 723–729 (1976).
146. Laing, W.A. & Christeller, J.T. Model for kinetics of activation and catalysis of ribulose 1,5-bisphosphate carboxylase. *Biochem. J.* **159**, 563–570 (1976).

References

147. Lorimer, G.H. & Miziorko, H.M. Carbamate formation on the epsilon-amino group of a lysyl residue as the basis for the activation of ribulosebiphosphate carboxylase by CO₂ and Mg²⁺. *Biochemistry* **19**, 5321–5328 (1980).
148. Donnelly, M.I., Stringer, C.D., & Hartman, F.C. Characterization of the activator site of *Rhodospirillum-rubrum* ribulosebiphosphate carboxylase oxygenase. *Biochemistry* **22**, 4346–4352 (1983).
149. Miziorko, H.M. Ribulose-1,5-bisphosphate carboxylase - evidence in support of the existence of distinct CO₂ activator and CO₂ substrate sites. *J. Biol. Chem.* **254**, 270–272 (1979).
150. Lorimer, G.H. Evidence for the existence of discrete activator and substrate sites for CO₂ on ribulose-1,5-bisphosphate carboxylase. *J. Biol. Chem.* **254**, 5599–5601 (1979).
151. Mueller-Cajar, O., Stotz, M., & Bracher, A. Maintaining photosynthetic CO₂ fixation via protein remodelling: the rubisco activases. *Photosynth. Res.* **119**, 191–201 (2014).
152. Andrews, T.J. & Ballment, B. Active-site carbamate formation and reaction-intermediate-analog binding by ribulosebiphosphate carboxylase/oxygenase in the absence of its small subunits. *Proc. Natl. Acad. Sci. United States Am. Sci.* **81**, 3660–3664 (1984).
153. Christeller, J.T. The effects of bivalent-cations on ribulose bisphosphate carboxylase-oxygenase. *Biochem. J.* **193**, 839–844 (1981).
154. Robison, P.D., Martin, M.N., & Tabita, F.R. Differential effects of metal-ions on *Rhodospirillum-rubrum* ribulosebiphosphate carboxylase-oxygenase and stoichiometric incorporation of HCO₃⁻ into a cobalt(III)-enzyme complex. *Biochemistry* **18**, 4453–4458 (1979).
155. Badger, M.R. & Lorimer, G.H. Interaction of sugar phosphates with the catalytic site of ribulose-1,5-bisphosphate carboxylase. *Biochemistry* **20**, 2219–2225 (1981).
156. Hatch, A.L. & Jensen, R.G. Regulation of ribulose-1,5-bisphosphate carboxylase from tobacco - changes in pH response and affinity for CO₂ and Mg²⁺ induced by chloroplast intermediates. *Arch. Biochem. Biophys.* **205**, 587–594 (1980).
157. Parry, M.A.J., Keys, A.J., Madgwick, P.J., Carmo-Silva, A.E., & Andralojc, P.J. Rubisco regulation: a role for inhibitors. *J. Exp. Bot.* **59**, 1569–1580 (2008).
158. Pearce, F.G. Catalytic by-product formation and ligand binding by ribulose bisphosphate carboxylases from different phylogenies. *Biochem. J.* **399**, 525–534 (2006).
159. Salvucci, M.E., Portis, A.R., & Ogren, W.L. A soluble chloroplast protein catalyzes ribulosebiphosphate carboxylase oxygenase activation in vivo. *Photosynth. Res.* **7**, 193–201 (1985).
160. Stotz, M. *et al.* Structure of green-type rubisco activase from tobacco. *Nat. Struct. Mol. Biol.* **18**, 1366–1370 (2011).
161. Gutteridge, S. *et al.* A nocturnal inhibitor of carboxylation in leaves. *Nature* **324**, 274–276 (1986).
162. Robinson, S.P. & Portis, A.R. Release of the nocturnal inhibitor, carboxyarabinitol-1-phosphate, from ribulose bisphosphate carboxylase oxygenase by rubisco activase. *Febs Lett.* **233**, 413–416 (1988).
163. Holbrook, G.P., Bowes, G., & Salvucci, M.E. Degradation of 2-carboxyarabinitol 1-phosphate by a specific chloroplast phosphatase. *Plant Physiol.* **90**, 673–678 (1989).

164. Heo, J.Y. & Holbrook, G.P. Regulation of 2-carboxy-d-arabinitol 1-phosphate phosphatase: activation by glutathione and interaction with thiol reagents. *Biochem. J.* **338**, 409–416 (1999).
165. Bracher, A., Sharma, A., Starling-Windhof, A., Hartl, F.U., & Hayer-Hartl, M. Degradation of potent rubisco inhibitor by selective sugar phosphatase. *Nat. Plants* **1**, (2015).
166. Pierce, J., Lorimer, G.H., & Reddy, G.S. Kinetic mechanism of ribulosebisphosphate carboxylase - evidence for an ordered, sequential reaction. *Biochemistry* **25**, 1636–1644 (1986).
167. Kannappan, B. & Gready, J.E. Redefinition of rubisco carboxylase reaction reveals origin of water for hydration and new roles for active-site residues. *J. Am. Chem. Soc.* **130**, 15063–15080 (2008).
168. Andrews, T.J. & Whitney, S.M. Manipulating ribulose biphosphate carboxylase/oxygenase in the chloroplasts of higher plants. *Arch. Biochem. Biophys.* **414**, 159–169 (2003).
169. Spreitzer, R.J. & Salvucci, M.E. Rubisco: structure, regulatory interactions, and possibilities for a better enzyme. *Annu. Rev. Plant Biol.* **53**, 449–475 (2002).
170. Parry, M.A.J., Andralojc, P.J., Mitchell, R.A.C., Madgwick, P.J., & Keys, A.J. Manipulation of rubisco: the amount, activity, function and regulation. *J. Exp. Bot.* **54**, 1321–1333 (2003).
171. Long, S.P., Zhu, X.G., Naidu, S.L., & Ort, D.R. Can improvement in photosynthesis increase crop yields? *Plant Cell Environ.* **29**, 315–330 (2006).
172. Whitney, S.M., Baldett, P., Hudson, G.S., & Andrews, T.J. Form I rubiscos from non-green algae are expressed abundantly but not assembled in tobacco chloroplasts. *Plant J.* **26**, 535–547 (2001).
173. Hartman, F.C. & Harpel, M.R. Structure, function, regulation, and assembly of d-ribulose-1,5-bisphosphatecarboxylase oxygenase. *Annu. Rev. Biochem.* **63**, 197–234 (1994).
174. Chua, N.H. & Schmidt, G.W. Post-translational transport into intact chloroplasts of a precursor to the small subunit of ribulose-1,5-bisphosphate carboxylase. *Proc. Natl. Acad. Sci. U. S. A.* **75**, 6110–6114 (1978).
175. Robinson, C. & Ellis, R.J. Transport of proteins into chloroplasts - partial-purification of a chloroplast protease involved in the processing of imported precursor polypeptides. *Eur. J. Biochem.* **142**, 337–342 (1984).
176. Tabita, F.R. Molecular and cellular regulation of autotrophic carbon-dioxide fixation in microorganisms. *Microbiol. Rev.* **52**, 155–189 (1988).
177. Hwang, S.R. & Tabita, F.R. Cotranscription, deduced primary structure, and expression of the chloroplast-encoded *rbcl* and *rbcs* genes of the marine diatom *Cylindrotheca* sp strain-N1. *J. Biol. Chem.* **266**, 6271–6279 (1991).
178. Smith, S.M. & Ellis, R.J. Processing of small subunit precursor of ribulose bisphosphate carboxylase and its assembly into whole enzyme are stromal events. *Nature* **278**, 662–664 (1979).
179. Ellis, R.J. & Van Der Vies, S.M. The rubisco subunit binding protein. *Photosynth. Res.* **16**, 101–115 (1988).
180. Gatenby, A.A., Lubben, T.H., Ahlquist, P., & Keegstra, K. Imported large subunits of ribulose bisphosphate carboxylase/oxygenase, but not imported beta-ATP synthase subunits, are assembled into holoenzyme in isolated chloroplasts. *EMBO J.* **7**, 1307–1314 (1988).

References

181. Barraclough, R. & Ellis, R.J. Protein synthesis in chloroplasts. IX. Assembly of newly-synthesized large subunits into ribulose biphosphate carboxylase in isolated intact pea chloroplasts. *Biochim. Biophys. Acta* **608**, 19–31 (1980).
182. Lubben, T.H., Donaldson, G.K., Viitanen, P. V, & Gatenby, A.A. Several proteins imported into chloroplasts form stable complexes with the GroEL-related chloroplast molecular chaperone. *Plant Cell* **1**, 1223–1230 (1989).
183. Zabaleta, E. *et al.* Antisense expression of chaperonin 60-beta in transgenic tobacco plants leads to abnormal phenotypes and altered distribution of photoassimilates. *Plant J.* **6**, 425–432 (1994).
184. Madueno, F., Napier, J.A., & Gray, J.C. Newly imported rieske iron-sulfur protein associates with both cpn60 and hsp70 in the chloroplast stroma. *Plant Cell* **5**, 1865–1876 (1993).
185. Tsugeki, R. & Nishimura, M. Interaction of homologs of hsp70 and cpn60 with ferredoxin-NADP+ reductase upon its import into chloroplasts. *Febs Lett.* **320**, 198–202 (1993).
186. Checa, S.K. & Viale, A.M. The 70-kda heat-shock protein DnaK chaperone system is required for the productive folding of ribulose-bisphosphate carboxylase subunits in *Escherichia coli*. *Eur. J. Biochem.* **248**, 848–855 (1997).
187. Andrews, T.J. & Lorimer, G.H. Catalytic properties of a hybrid between cyanobacterial large subunits and higher-plant small subunits of ribulose biphosphate carboxylase-oxygenase. *J. Biol. Chem.* **260**, 4632–4636 (1985).
188. Roy, H., Bloom, M., Milos, P., & Monroe, M. Studies on the assembly of large subunits of ribulose biphosphate carboxylase in isolated pea-chloroplasts. *J. Cell Biol.* **94**, 20–27 (1982).
189. Bloom, M.V, Milos, P., & Roy, H. Light-dependent assembly of ribulose-1,5-bisphosphate carboxylase. *Proc. Natl. Acad. Sci. United States Am. Sci.* **80**, 1013–1017 (1983).
190. Goloubinoff, P., Gatenby, A.A., & Lorimer, G.H. GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose biphosphate carboxylase oligomers in *Escherichia coli*. *Nature* **337**, 44–47 (1989).
191. Goloubinoff, P., Christeller, J.T., Gatenby, A.A., & Lorimer, G.H. Reconstitution of active dimeric ribulose biphosphate carboxylase from an unfolded state depends on 2 chaperonin proteins and Mg-ATP. *Nature* **342**, 884–889 (1989).
192. Viitanen, P. V *et al.* Chaperonin-facilitated refolding of ribulosebisphosphate carboxylase and ATP hydrolysis by chaperonin-60 (GroEL) are K⁺ dependent. *Biochemistry* **29**, 5665–5671 (1990).
193. Baneyx, F. *et al.* Spinach chloroplast cpn21 co-chaperonin possesses 2 functional domains fused together in a toroidal structure and exhibits nucleotide-dependent binding to plastid chaperonin-60. *J. Biol. Chem.* **270**, 10695–10702 (1995).
194. Gatenby, A.A. Synthesis and assembly of bacterial and higher-plant rubisco subunits in *Escherichia coli*. *Photosynth. Res.* **17**, 145–157 (1988).
195. Cloney, L.P., Bekkaoui, D.R., Wood, M.G., & Hemmingsen, S.M. Assessment of plant chaperonin-60 gene-function in *Escherichia coli*. *J. Biol. Chem.* **267**, 23333–23336 (1992).
196. Cloney, L.P., Bekkaoui, D.R., & Hemmingsen, S.M. Coexpression of plastid chaperonin genes and a synthetic plant rubisco operon in *Escherichia coli*. *Plant Mol. Biol.* **23**, 1285–1290 (1993).
197. Larimer, F.W. & Soper, T.S. Overproduction of *Anabaena*-7120 ribulose-bisphosphate carboxylase oxygenase in *Escherichia coli*. *Gene* **126**, 85–92 (1993).

198. Emlyn-Jones, D., Woodger, F.J., Price, G.D., & Whitney, S.M. RbcX can function as a rubisco chaperonin, but is non-essential in *Synechococcus* PCC7942. *Plant Cell Physiol.* **47**, 1630–1640 (2006).
199. Li, L.A. & Tabita, F.R. Maximum activity of recombinant ribulose 1,5-bisphosphate carboxylase/oxygenase of *Anabaena* sp. strain CA requires the product of the RbcX gene. *J. Bacteriol.* **179**, 3793–3796 (1997).
200. Onizuka, T. *et al.* The rbcx gene product promotes the production and assembly of ribulose-1,5-bisphosphate carboxylase/oxygenase of *Synechococcus* sp. PCC7002 in *Escherichia coli*. *Plant Cell Physiol.* **45**, 1390–1395 (2004).
201. Saschenbrecker, S. *et al.* Structure and function of RbcX, an assembly chaperone for hexadecameric rubisco. *Cell* **129**, 1189–1200 (2007).
202. Andrews, T.J. Catalysis by cyanobacterial ribulose-bisphosphate carboxylase large subunits in the complete absence of small subunits. *J. Biol. Chem.* **263**, 12213–12219 (1988).
203. Liu, C. *et al.* Coupled chaperone action in folding and assembly of hexadecameric rubisco. *Nature* **463**, 197–202 (2010).
204. Bracher, A., Starling-Windhof, A., Hartl, F.U., & Hayer-Hartl, M. Crystal structure of a chaperone-bound assembly intermediate of form I rubisco. *Nat. Struct. Mol. Biol.* **18**, 875–880 (2011).
205. Hauser, T. *et al.* Structure and mechanism of the rubisco-assembly chaperone Raf1. *Nat. Struct. Mol. Biol.* **22**, 720–728 (2015).
206. Joshi, J., Mueller-Cajar, O., Tsai, Y.C., Hartl, F.U., & Hayer-Hartl, M. Role of small subunit in mediating assembly of red-type form I rubisco. *J Biol Chem* **290**, 1066–1074 (2015).
207. Alonso, H., Blayney, M.J., Beck, J.L., & Whitney, S.M. Substrate-induced assembly of *Methanococcoides burtonii* d-ribulose-1,5-bisphosphate carboxylase/oxygenase dimers into decamers. *J. Biol. Chem.* **284**, 33876–33882 (2009).
208. Van Lun, M., van der Spoel, D., & Andersson, I. Subunit interface dynamics in hexadecameric rubisco. *J Mol Biol* **411**, 1083–1098 (2011).
209. Feiz, L. *et al.* Ribulose-1,5-bis-phosphate carboxylase/oxygenase accumulation factor1 is required for holoenzyme assembly in maize. *Plant Cell* **24**, 3435–3446 (2012).
210. Feiz, L. *et al.* A protein with an inactive pterin-4a-carbinolamine dehydratase domain is required for rubisco biogenesis in plants. *Plant J.* **80**, 862–869 (2014).
211. Naponelli, V. *et al.* Phylogenomic and functional analysis of pterin-4a-carbinolamine dehydratase family (COG2154) proteins in plants and microorganisms. *Plant Physiol.* **146**, 1515–1527 (2008).
212. Hauser, T., Popilka, L., Hartl F.U., & Hayer-Hartl, M. Role of auxiliary proteins in rubisco biogenesis and function. *Nat. Plants* **1**, (2015).
213. Viitanen, P. V *et al.* Functional characterization of the higher plant chloroplast chaperonins. *J Biol Chem* **270**, 18158–18164 (1995).
214. Kolesinski, P., Piechota, J., & Szczepaniak, A. Initial characteristics of RbcX proteins from *Arabidopsis thaliana*. *Plant Mol. Biol.* **77**, 447–459 (2011).
215. Kolesinski, P. *et al.* Insights into eukaryotic rubisco assembly - crystal structures of RbcX chaperones from *Arabidopsis thaliana*. *Biochim. Biophys. Acta* **1830**, 2899–2906 (2013).

References

216. Bracher, A., Hauser, T., Liu, C., Hartl, F.U., & Hayer-Hartl, M. Structural analysis of the rubisco-assembly chaperone rbcx-II from *Chlamydomonas reinhardtii*. *PLoS One* **10**, (2015).
217. Rohr, J., Sarkar, N., Balenger, S., Jeong, B.R., & Cerutti, H. Tandem inverted repeat system for selection of effective transgenic RNAi strains in *Chlamydomonas*. *Plant J.* **40**, 611–621 (2004).
218. Kolesinski, P., Belusiak, I., Czarnocki-Cieciura, M., & Szczepaniak, A. Rubisco accumulation factor 1 from *Thermosynechococcus elongatus* participates in the final stages of ribulose-1,5-bisphosphate carboxylase/oxygenase assembly in *Escherichia coli* cells and in vitro. *FEBS J* **281**, 3920–3932 (2014).
219. Tarnawski, M., Gubernator, B., Kolesinski, P., & Szczepaniak, A. Heterologous expression and initial characterization of recombinant RbcX protein from *Thermosynechococcus elongatus* BP-1 and the role of RbcX in rubisco assembly. *Acta Biochim Pol* **55**, 777–785 (2008).
220. Whitney, S.M., Birch, R., Kelso, C., Beck, J.L., & Kapralov, M. V. Improving recombinant rubisco biogenesis, plant photosynthesis and growth by coexpressing its ancillary Raf1 chaperone. *Proc Natl Acad Sci U S A* **112**, 3564–3569 (2015).
221. Chari, A. & Fischer, U. Cellular strategies for the assembly of molecular machines. *Trends Biochem Sci* **35**, 676–683 (2010).
222. Ellis, R.J. Assembly chaperones: a perspective. *Philos Trans R Soc L. B Biol Sci* **368**, 20110398 (2013).
223. Tanaka, S., Sawaya, M.R., Kerfeld, C.A., & Yeates, T.O. Structure of the rubisco chaperone RbcX from *Synechocystis* sp. PCC6803. *Acta Crystallogr D Biol Crystallogr* **63**, 1109–1112 (2007).
224. Genkov, T., Meyer, M., Griffiths, H., & Spreitzer, R.J. Functional hybrid rubisco enzymes with plant small subunits and algal large subunits: engineered rbcS cDNA for expression in *Chlamydomonas*. *J Biol Chem* **285**, 19833–19841 (2010).
225. Hohmann-Marriott, M.F. & Blankenship, R.E. Evolution of photosynthesis. *Annu. Rev. Plant Biol.* **62**, 515–548 (2011).
226. Georgescauld, F. *et al.* GroEL/ES chaperonin modulates the mechanism and accelerates the rate of tim-barrel domain folding. *Cell* **157**, 922–934 (2014).
227. Chaudhuri, T.K. & Gupta, P. Factors governing the substrate recognition by GroEL chaperone: a sequence correlation approach. *Cell Stress Chaperones* **10**, 24–36 (2005).
228. Schottkowski, M. *et al.* Biogenic membranes of the chloroplast in *Chlamydomonas reinhardtii*. *Proc Natl Acad Sci U S A* **109**, 19286–19291 (2012).
229. Wheatley, N.M., Sundberg, C.D., Gidaniyan, S.D., Cascio, D., & Yeates, T.O. Structure and identification of a pterin dehydratase-like protein as a ribulose-bisphosphate carboxylase/oxygenase (rubisco) assembly factor in the alpha-carboxysome. *J Biol Chem* **289**, 7973–7981 (2014).
230. Brutnell, T.P., Sawers, R.J., Mant, A., & Langdale, J.A. Bundle sheath defective2, a novel protein required for post-translational regulation of the *rbcl* gene of maize. *Plant Cell* **11**, 849–864 (1999).
231. Doron, L., Segal, N., Gibori, H., & Shapira, M. The *bsd2* ortholog in *Chlamydomonas reinhardtii* is a polysome-associated chaperone that co-migrates on sucrose gradients with the *rbcl* transcript encoding the rubisco large subunit. *Plant J.* **80**, 345–355 (2014).

232. Kruger, E.L. & Volin, J.C. Reexamining the empirical relation between plant growth and leaf photosynthesis. *Funct. Plant Biol.* **33**, 421–429 (2006).
233. Sambrook, J., Fritsch, E. & Maniatis, T. *Molecular Cloning: A Laboratory Manual. Second Edition.* (Cold Spring Harbor Press, 1989).
234. Weiner, M.P. *et al.* Site-directed mutagenesis of double-stranded DNA by the polymerase chain-reaction. *Gene* **151**, 119–123 (1994).
235. Baker, R.T. *et al.* Using deubiquitylating enzymes as research tools. *Ubiquitin Protein Degrad. Part A* **398**, 540–554 (2005).
236. Stoscheck, C.M. Quantitation of protein. *Methods Enzymol.* **182**, 50–68 (1990).
237. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. & Struhl, K. *Current Protocols in Molecular Biology.* (John Wiley & Sons, 1997).
238. Bradford, M.M. Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).
239. Laemmli, U.K. Cleavage of structural proteins during assembly of head of bacteriophage-T4. *Nature* **227**, 680–& (1970).
240. Towbin, H., Staehelin, T., & Gordon, J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets - procedure and some applications. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 4350–4354 (1979).
241. Coligan, J.E., Dunn B.M., Speicher, D.W. & Wingfield, P.T. Proteins: structure, function, and genetics. In *Current Protocols in Protein Science* **24**, (John Wiley & Sons, Inc., New York, 1995).
242. Somerville, C.R., McIntosh, L., Fitchen, J., & Gurevitz, M. The cloning and expression in *Escherichia coli* of RuBP carboxylase oxygenase large subunit genes. *Methods Enzymol.* **118**, 419–433 (1986).
243. Harris, E.H. *The Chlamydomonas Source Book.* (Academic Press, San Diego, 2009).
244. Neupert, J., Karcher, D., & Bock, R. Generation of *Chlamydomonas* strains that efficiently express nuclear transgenes. *Plant J.* **57**, 1140–1150 (2009).
245. Kindle, K.L. High-frequency nuclear transformation of *Chlamydomonas reinhardtii*. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 1228–1232 (1990).
246. Palombella, A.L. & Dutcher, S.K. Identification of the gene encoding the tryptophan synthase beta-subunit from *Chlamydomonas reinhardtii*. *Plant Physiol.* **117**, 455–464 (1998).

8. Appendices

8.1 Oligonucleotides

Tab.A1: Sequences of oligonucleotides used in this study

| name | Sequence (5'-3') |
|---------------------|--|
| AtRbcXI_FW | CTCCGCGGTGGAGAGGATGTTGCTGGTAATTACGACGAT ACC |
| AtRbcXI_RV | GTCAAGCTTTTACCTTGAGTTTGTGTCATCGG |
| AtRbcXII_FW | CTCCGCGGTGGAAAGATGTATGTTCCCGGCTTTGGAGAA GC |
| AtRbcXII_RV | GTCAAGCTTTCACTTATCAGAATCGGTTTCGACATGG |
| AtRbcXI_R151A+ | GGAGCTTGCAGAAGCAGTGATGATCACG |
| AtRbcXI_R151A- | CGTGATCATCACTGCTTCTGCAAGCTCC |
| AtRbcXII_R123A+ | GAACCTAGCCCTTGCAATTTTAGAGGTACG |
| AtRbcXII_R123A- | CGTACCTCTAAAATTGCAAGGGCTAAGTTC |
| CrRbcX570EcoRI | TTCGAATTCATGAGCCTCAGCGCAAGAGC |
| CrRbcX570BamHI | TTCGGATCCTCACGCGGCACCCTTGCCGG |
| CrRbcX370BamHIEcoRI | TTCGGATCCGAATTCATGAGCCTCAGCGCAAGAGC |
| CrRbcX370BglII | TTCAGATCTCGCGCACCTCCATCAGGCGC |
| CRrbcXRNAiFW | CCTGGACGAGTTTGAGTGGGGAAAG |
| CrRbcXRNAiRV | CTTGCCGGGGCCGTCCAGGTCCTCC |
| CrGAPDHFV | GGCAAGATTAAGATCGGCATCAAC |
| CrGAPDHRV | GTGGTCATCAGGCCCTCCTTG |
| CrRbcX-IIbFW | CTCCGCGGTGGTATGTACGTCCCATCGGACAGCTTC |
| CrRbcX-IIbRV | CGTGAGCTCTCATGCCTTCTCCCCATCCGTCG |
| CrRbcX-IIb(-SacII)+ | GAGCCTACGACTCACAGGCTGCAGGTGACCTGCACCACT TCCTC |

| | |
|---------------------|--|
| CrRbcX-IIb(-SacII)- | GAGGAAGTGGTGCAGGTCACCTGCAGCCTGTGAGTCGTA GGCTC |
| AtRaf2FW | CTCCGCGGTGGTTCAATCCTTAAAGATTTTCTTGGTGACT TCGGTG |
| AtRaf2RV | CGTGAGCTCTCACGCCCAAGCTCTTTTCCTAG |
| Syn7002Raf2FW | CTCCGCGGTGGTATGGCAACCCGATTGACCGAC |
| Syn7002Raf2RV | CGTGAGCTCTTACAGTTGATCGAAAGTTC |

8.2 Plasmids

Tab.A2: List of plasmids constructed in this study

| name | genes |
|--------------------|---|
| pHue-AtRbcXI | RbcXI from <i>Arabidopsis thaliana</i> |
| pHue-AtRbcXIR151A | RbcXI from <i>Arabidopsis thaliana</i> harbouring R151A mutation |
| pHue-AtRbcXII | RbcXII from <i>Arabidopsis thaliana</i> |
| pHue-AtRbcXIIR123A | RbcXII from <i>Arabidopsis thaliana</i> harbouring R123A mutation |
| NE537-CrRbcXRNAi | RbcX-IIa from <i>Chlamydomonas reinhardtii</i> |
| pHue-CrRbcX-IIb | RbcX-IIb from <i>Chlamydomonas reinhardtii</i> |
| pHue-AtRaf2 | Raf2 from <i>Arabidopsis thaliana</i> (aa 51-220, at5g51110) |
| pHue-Syn7002Raf2 | Raf2 from <i>Synechococcus</i> sp. PCC7002 |

8.3 Abbreviations

| | |
|-----------------|--|
| Å | Angström |
| AA | acrylamide |
| aa | amino acid |
| ADP | adenosine 5'-diphosphate |
| AnaCA | <i>Anabaena</i> CA |
| APS | ammonium peroxodisulfate |
| At | <i>Arabidopsis thaliana</i> |
| ATP | adenosine 5'-triphosphate |
| BSA | bovine serum albumin |
| °C | degree Celsius |
| CABP | 2-carboxyarabinitol biphosphate |
| CA1P | carboxyarabibitol 1-phosphate |
| CAM | Crassulacean Acid Metabolism |
| CBB | Calvin-Benson-Bassham |
| CO ₂ | carbondioxide |
| Cpn | chaperonin |
| Cr | <i>Chlamydomonas reinhardtii</i> |
| DNA | deoxyribonucleic acid |
| DnaJ | bacterial Hsp40 chaperone |
| DnaK | bacterial Hsp70 chaperone |
| dNTP | 2'-desoxyribonucleotide-triphosphate |
| DTT | dithiothreitol |
| <i>E. coli</i> | <i>Escherichia coli</i> |
| EDTA | ethylenediaminetetraacetic acid |
| g | acceleration of gravity, 9.81 m/s ² |
| GroEL | bacterial Hsp60 chaperonin |
| GroES | bacterial Hsp10 co-chaperonin |
| G3P | glyceraldehyde phosphate |
| h | hour |
| HMW | high molecular weight |
| HRP | horseradish peroxidase |
| Hsp | heat shock protein |
| IP | immunoprecipitation |

| | |
|------------------|---|
| IPTG | isopropyl- β -D-1-thiogalactopyranoside |
| KABP | 3-ketoarabinitol-1,5-bisphosphate |
| k_{cat}^C | CO ₂ -saturated carboxylase velocity |
| k_{cat}^O | O ₂ -saturated oxygenase velocity |
| K_C | Michaelis-Menten constant for carboxylation |
| K_O | Michaelis-Menten constant for oxygenation |
| kDa | kilodalton |
| LB | Luria Bertani |
| Mg ²⁺ | magnesium |
| min. | minutes |
| MS | Mass Spectrometry |
| MW | molecular weight |
| NAC | nascent chain-associated complex |
| NADPH | β -nicotinamide adenine dinucleotide 2'-phosphate |
| NBD | nucleotide-binding domain |
| NEF | nucleotide-exchange factor |
| O ₂ | oxygen |
| OD | optical density |
| o/n | overnight |
| PAGE | polyacrylamide gel electrophoresis |
| PCR | polymerase chain reaction |
| PDB | protein data bank |
| PFD | prefoldin |
| Pfu | <i>Pyrococcus furiosus</i> |
| 3PG | 3-phosphoglycerate |
| PEP | phosphoenolpyruvate |
| PDBP | 2,3-pentodiulose-1,5-bisphosphate |
| pmf | proton motif force |
| PMSF | phenyl-methyl-sulfonyl fluoride |
| RAC | ribosome-associated complex |
| Raf1 | Rubisco accumulation factor 1 |
| Raf2 | Rubisco accumulation factor 2 |
| RbcL | Rubisco large subunit |
| RbcS | Rubisco small subunit |

Appendices

| | |
|----------|--|
| Rca | Rubisco activase |
| RLP | Rubisco-like protein |
| Rubisco | Ribulose-1,5-bisphosphate carboxylase/oxygenase |
| RuBP | Ribulose-1,5-bisphosphate |
| RNA | ribonucleic acid |
| rpm | revolutions per minute |
| SBD | substrate-binding domain |
| SEC-MALS | size-exclusion chromatography – multi-angle light scattering |
| SDS | sodiumdodecylsulfate |
| sp. | species |
| Syn | <i>Synechococcus</i> |
| TEMED | <i>N,N,N',N'</i> -tetramethylethylenediamine |
| TF | trigger factor |
| THF | tetrahydrofolate |
| TRiC | TCP-1 containing ring complex |
| Tris | trishydroxymethylaminomethan |
| v/v | volume per volume |
| w/v | weight per volume |
| XuBP | xylulose-1,5-bisphosphate |

8.4 Curriculum Vitae

Personal Details

Name Thomas Hauser
Date of birth 1982/12/14
Place of birth Tübingen, Germany

Higher Education and International Work Experience

08/2009 – 11/2015 **Dissertation**
Max-Planck-Institute of Biochemistry, Martinsried, Germany
Department of Cellular Biochemistry, group of Prof. Dr. F. Ulrich Hartl
Thesis title: Structural and Functional Characterization of Rubisco Assembly Chaperones

10/2003 – 05/2009 **Subject of Study Biotechnology, academic title: “Diplom Biologe technisch orientiert”, equivalent to MSc.**
University Stuttgart, Germany
Main Subjects: Biochemical Engineering, Biochemistry
Secondary Subjects: Bioinformatics, Microbiology

08/2008 – 03/2009 **Diploma thesis**
Max-Planck-Institute of Biochemistry, Martinsried, Germany
Department of Cellular Biochemistry, group of Prof. Dr. F. Ulrich Hartl
Thesis title: Folding and Assembly of Form I Rubisco

05/2007 – 10/2007 **Thesis Research Project**
Yale University, New Haven, USA
Department of Molecular Biophysics and Biochemistry, group of Prof. Dr. Mark Hochstrasser
Thesis title: ERAD in *S.cerevisiae*

Publications

2015 Thomas Hauser, Javaid Y. Bhat, Goran Miličić, Petra Wendler, F. Ulrich Hartl, Andreas Bracher and Manajit Hayer-Hartl: Structure and mechanism of the Rubisco assembly chaperone Raf1, *Nature Structure and Molecular Biology*, 22, 720-728.

Andreas Bracher*, Thomas Hauser*, Cuimin Liu, F. Ulrich Hartl and Manajit Hayer-Hartl: Structural analysis of the Rubisco-assembly chaperone RbcX-II from *Chlamydomonas reinhardtii*, *PLOS ONE*.

Thomas Hauser, Leonhard Popilka, Ulrich Hartl and Manajit Hayer-Hartl: Role of auxiliary proteins in Rubisco biogenesis and Function, *Nature Plants*.

2008 Franz Brümmer, Martin Pfannkuchen, Alexander Baltz, Thomas Hauser and Vera Thiel: Light inside sponges, *Journal of Experimental Marine Biology and Ecology*, 367(2), 61-64.